5-2013

# Evolutionary and Molecular Analysis of Conserved Vertebrate Immunity to Fungi

Erin Carter
*University of Maine - Main*

# EVOLUTIONARY AND MOLECULAR ANALYSIS OF CONSERVED VERTEBRATE IMMUNITY TO FUNGI

by

Erin Victoria Carter

A Thesis Submitted in Partial Fulfillment
of the Requirements for a Degree with Honors
(Biochemistry, Molecular and Cellular Biology, Microbiology)

The Honors College

University of Maine

May 2013

Advisory Committee:

Robert Wheeler, Assistant Professor of Microbiology, Advisor
Julie Gosse, Assistant Professor of Biochemistry
Robert Gundersen, Associate Professor of Biochemistry and Chair of Molecular and Biomedical Sciences
Keith Hutchison, Professor of Biochemistry and Molecular Biology
Mimi Killinger, Adjunct Assistant Professor in Honors (History)

# ABSTRACT

The innate immune system is highly conserved amongst all multicellular organisms. Yet a constant battle exists between host cells and pathogens due to the rapid evolution of immune system components. Functional genomics and *in silico* methods can be employed to elucidate the evolutionary patterns of vertebrate immunity to pathogenic fungi such as *Candida albicans*, an opportunistic fungal pathogen that can cause lethal candidiasis in the immunocompromised. Mammals such as humans and mice possess conserved C-type lectin receptors that recognize the *C. albicans* cell wall. However, these receptors have not been identified in fish. Here I describe how we identified potential zebrafish fungal recognition receptors *in silico* to elucidate the evolution of vertebrate immunity to fungi and to integrate cost-effective zebrafish into candidiasis research. Phylogenetic and synteny analyses identified three potential receptors with conserved motifs for fungal recognition. Cell lines secreting soluble versions of these potential receptors were generated, and the proteins were purified. These receptors are currently being analyzed for their microbial recognition characteristics through immunofluorescence microscopy. Determining the specificity of these proteins may enhance our understanding of how the innate immune system evolved in lower and higher vertebrates. Furthermore, understanding such dynamics is an initial step toward developing novel anti-fungal therapeutics for commercially valuable fish and uncovering the fundamental mechanisms of immunity to fungi.

ACKNOWLEDGEMENTS

There are many talented individuals who I want to acknowledge regarding my Honors Thesis. Without their input, support, and patience, this thesis would probably not exist.

First and foremost, I am grateful for all members of the Wheeler Lab, past and present. Their insight and support has been valuable during these past three years. In particular, I would like to thank Dr. Remi Gratacap for preparing the zebrafish cDNA library and for guidance at the bench, Dr. Kimberly Brothers-Walls for guidance at the bench, Riju Shrestha for assistance while cloning, Nadine Nicke for establishing the cell lines, and Alex Hopke for establishing the cell lines and for purifying the zSCLRA-Fc fusion protein.

I also wish to express my gratitude toward several individuals outside of the Wheeler Lab. I owe a debt of thanks to Garth Holman from Dr. Campbell's Lab for his technical assistance regarding phylogenetic software and multiple alignments. His advice has been extremely valuable in both generating multiple alignments and phylogenetic trees. I would also like to thank Cathy Hopper for providing me with a strain of *Bacillus subtilis* to utilize for future experiments that test the binding properties of the novel fusion proteins. I also wish to thank each of my Thesis Committee Members for agreeing to serve on my committee and for their insight on various aspects of my thesis.

Additionally, I want to acknowledge every brave soul who has listened to me discuss my thesis work over the years in some context. Their patience, suggestions, and support have been incredibly helpful.

And finally, I owe a special thank you to Dr. Rob Wheeler for being a terrific mentor and Thesis advisor who has helped me at many points along the Thesis writing journey!

TABLE OF CONTENTS

# INTRODUCTION

## i. Overview

Fungal pathogens have plagued all known multicellular organisms for centuries, including humans. With incidents of these potentially dangerous pathogens on the rise, it would be useful to understand host-pathogen interactions in regards to fungi and how this immune response has evolved from fish to mammals. We are particularly interested in the conservation of innate immune receptors specific for fungi. In this Introduction, I will first discuss how comparative functional genomics can be used to address this question. Species comparisons can yield valuable information about conserved aspects of species-specific innate immune response. This can be done using phylogenetics, synteny, and sequence homology. Next, I will review the relevance of mammalian innate immune responses to the opportunistic fungal pathogen *Candida albicans*, as C-type lectin pattern recognition receptors recognize fungi and trigger an immune response to the pathogen. I will then explore the emerging role of zebrafish as a model organism and how critical components of the innate and adaptive immune system are conserved between zebrafish and humans. Although aspects of the zebrafish genome are highly conserved amongst mammalian genomes, fungal recognition receptors have yet to be identified in the zebrafish. I will discuss how many aspects of C-type lectin pattern recognition specificity are currently elusive and how comparative functional genomics can be employed to address this area of research.

ii. Cross-Kingdom Significance of Fungal Pathogens

Fungi are ubiquitous in nature and cause disease in all known plants and animals (12). Pathogenic fungi have notoriously infected plants in the past, yet in recent years, novel fungal pathogens have affected animal populations, including bony fish, bats, and humans (13). While many fungal pathogens pose little threat to overall human health, some are invasive and opportunistic with high mortality rates (8). This universal battle against fungi could be helped by a better understanding of how hosts recognize fungi and how this recognition has evolved from fish to mammals. Because microbial recognition by innate immune receptors is critical for mediating immunity, it is therefore important to understand how immune system receptors against fungi have evolved among vertebrates.

iii. Employment of Bioinformatics to Elucidate Protein Evolution

To study conserved innate immune responses between zebrafish and mammals, we focused on the identification of pattern recognition receptors (PRRs). PRRs recognize pathogen-associated molecular patterns (PAMPs) found on the surface of all pathogens. While the signaling pathways downstream of these receptors are highly conserved, there is little conservation between the receptors themselves among all vertebrates. This presents a complex problem in regards to conservation of innate immune recognition amongst multicellular organisms. Comparative functional genomics can be employed to address this issue. Bioinformatics and recombinant DNA methods can be used to compare genetic functions between different genes or species. This is particularly useful when analyzing the complex data associated with modern evolutionary biology studies. Evolutionary biology examines how organisms and their components evolved over time.

In collaboration, evolutionary biology and functional genomics can be applied to solve a variety of complex problems relating to the evolutionary relationships of a vast array of organisms, particularly in identifying sequences based on high levels of evolutionary conservation (31). It is especially of interest to identify how proteins such as the receptors associated with fungal immunity have evolved.

In terms of protein evolution, it has been shown that extracellular components evolve at a faster rate than intracellular components. Extracellular domains of proteins and protein kinases evolved faster than intracellular protein domains, as these domains are more divergent (29). Additionally, extracellular protein evolution has been shown to be independent of protein-protein interactions, expression levels, and immune function (22). Location was considered to be the most important factor concerning evolution rate (22); however, genetic function has had some effect on evolutionary rates. Transcriptional regulators evolve slower in protein and amino acid sequence, but faster in terms of expression levels (27). These are often located inside the cell and are considered to be essential. Various enzymes, receptors, transporters, and channels evolve faster in amino acid sequence (27), as these are less essential from cell to cell and are frequently extracellular. Since immune receptors are more often than not extracellular, it is likely that innate immune receptors evolve quickly. Such protein evolution can be studied using *in silico* methods such as phylogenetics and synteny.

*a. Phylogenetics in Theory and Practice*

Phylogenetics is employed to study the evolutionary relationships between organisms or sequences (17) and to predict classifications of genes, proteins, cellular

components, and organisms. A phylogeny can be of important use when studying the evolutionary relationships or relatedness of homologues, orthologues, and paralogues. In particular, phylogenetics analyzes evolutionary distances between DNA, RNA, and protein sequences and visualizes these distances in the form of a tree (17). A phylogenetic tree consists of several parts, including intermodal branches that connect various nodes (3). A node is representative of a differentiation or speciation event, and each terminal branch is representative of a species or sequence (3).

Protein phylogenies are currently constructed using computer algorithms and statistical models (17), as the data are simply too complex for the mind to solve alone with great efficiency. Prior to tree construction, sequences of interest must be aligned in order to maximize the accuracy of the tree (17); this is often done through the use of ClustalW, which has since been updated to ClustalOmega. At times, a statistical model is necessary prior to constructing phylogenies in order to see which statistical model yields the best phylogenetic tree (3, 17). Finally, an appropriate phylogenetic method must be selected. This determines exactly how the tree is constructed. Trees can be constructed using phylogenetic software. Currently, MEGA5 is the most powerful and efficient computer algorithm for constructing phylogenies, as it can align sequences, perform statistical model tests, and construct maximum parsimony, maximum likelihood, and neighbor-joining trees with high efficiency (44).

Maximum parsimony is the earliest method of phylogenetic construction, working on the basis of minimum evolution to find the simplest phylogenetic topology out of all hypothetical trees (3, 17). This algorithm examines every possible tree topology of a given alignment through ancestral sequences associated with each node of the tree (17).

Despite its simplistic model, maximum parsimony is not an efficient *in silico* method for our questions regarding innate immunity. This is due to the high mutation rates and different tertiary structural constraints. Thus, this method is not ideal for evolutionary analysis of fast-evolving immune receptors and is best used for closely related sequences (17).

Maximum likelihood is a newer phylogenetic method that utilizes an assumed probability model of evolutionary change to select the most likely phylogenetic topology (3, 25). Before constructing a maximum likelihood tree, a statistical model test must be performed to determine which model best determines the likelihood of mutations for a given set of data. Based on the statistical model, a maximum likelihood algorithm examines all possible tree topologies. This method has been proposed to be ideal for complex phylogenies, especially when comparing nucleotide sequences (17). Despite this, maximum likelihood methods may not be the best method for examining fast-evolving proteins, as the number of mutations may be so large that the data are inconsistent.

A third phylogenetic method, neighbor-joining, is growing in popularity. Neighbor-joining is based on evolutionary distances and clusters data from the bottom of the phylogeny to the top (26, 37). Unlike maximum parsimony and maximum likelihood trees, a neighbor-joining tree is unrooted and does not always examine every possible topology (37). Therefore, the simplest tree is not always generated. However, this method does provide branch lengths of the tree to estimate evolutionary distances (37), further differentiating it from other methods of tree construction. Neighbor-joining can be constructed using a Poisson distribution statistical model (26, 44) and does not require a

preliminary statistical model test. For most phylogenetic inferences, including those concerning fast-evolving proteins, neighbor-joining trees are highly desirable (17).

*b. Synteny in Theory and Practice*

Synteny is another important *in silico* concept to consider when studying genomic conservation and evolution, as it provides information concerning how genome structure has changed over evolutionary time. Synteny refers to the conservation of genomic location, especially in regard to genes that lie on the same chromosome (15). By performing a synteny analysis across several species, one can determine which genes have been highly conserved throughout evolutionary history. Furthermore, one can determine if genomic location has also been conserved. This technique is potentially valuable in identifying orthologues, or proteins that evolved from one common ancestor (11). Synteny can also be employed to identify paralgoues, which are proteins that emerged from gene duplication events (10).

It is generally accepted that there are four levels of genomic conservation: conserved synteny, conserved gene order, conserved orientation, and conserved block (11). Conserved synteny is the lowest level of genomic conservation (11) and simply refers to the conservation of genes on a particular chromosome between at least two species. A conserved block is considered to be the highest level of genomic conservation. In conserved blocks, gene order and orientation are conserved and the block is not interrupted by any additional genes (11). Therefore, species that possess conserved blocks in their genomes are believed to share higher levels of genomic conservation than those that possess basic conserved synteny.

In recent years, several high-throughput synteny models have been developed to assess levels of genomic conservation between organisms. Genome Rearrangements in Man and Mouse (GRIMM) was developed in 2002 to detect synteny blocks that are homologous across multiple genomes and to detect genome rearrangements (45). GRIMM relies on various algorithms to compute genomic distances. However, it is specifically adapted for analyzing two genomes only and would not be of use to researchers interested in synteny among multiple genomes. Multiple Genome Rearrangement (MGR) is an additional tool that reconstructs various rearrangement scenarios across multiple genomes on the basis of phylogeny (7). The algorithms behind both GRIMM and MGR are complex and both programs require Perl and Linux, indicating that they would require programming expertise for basic synteny analysis among zebrafish, humans, and mice.

Julian Catchen and his colleagues developed the Synteny Database in recent years. This database is an automated system that identifies conserved syntenic regions of genomes that uses an outgroup from a non-duplicated genome to help identify orthologues, paralogues, translocations, and inversions (10). The server is versatile in that it can examine data from fully annotated and partially annotated genomes (10). Thus, a synteny analysis can easily be performed between zebrafish and mammals. Like GRIMM, the Synteny Database can only perform synteny between two genomes at a time, which hinders one from directly performing a synteny analysis between three or more organisms.

A synteny analysis can also be performed manually if relatively simple syntenic relationships are being analyzed. Genomic data, including the arrangement of genes on

chromosomes, is available for most of the identified human, mouse, and zebrafish genes through the National Center of Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/) and Ensembl (http://useast.ensembl.org/index.html) databases. This information can be collected and arranged into a spatial representation without any restrictions that may be evident while using high throughput synteny methods.

iv. The Immune System and Candidemia

Functional genomics can be employed to elucidate the evolutionary and molecular mechanisms of conserved vertebrate immunity to fungi. In healthy individuals, the immune system controls disease by recognizing and distinguishing foreign substances from self-antigens. The innate immune system is the first line of defense against foreign substances. Innate immune components include epithelial barriers, commensal organisms, soluble factors, phagocytes, and pattern recognition receptors (24), and are highly conserved among all multicellular organisms. Some innate immune components collaborate with the adaptive immune system, which is less conserved and related to long-term defense against foreign antigens. An adaptive immune response eventually leads to specific cellular and humoral immunity directed at foreign antigens, thus inducing memory.

A current problem in modern infectious disease is the rise in opportunistic fungal pathogens. Although common, these organisms are understudied in comparison to viruses and bacteria (8). These pathogens are of little harm to individuals with healthy immune systems, and instead have adapted to live as commensal organisms. However,

HIV/AIDS, anti-cancer chemotherapies, organ transplants, and corticosteroids destroy vital components of the immune system, leaving individuals more susceptible to normally non-pathogenic microbes. *Candida albicans* is one of the most pervasive opportunistic fungal pathogens. This dimorphic, commensal fungus causes various infections in the immunocompromised, ranging from oral thrush to candidemia, a potentially lethal systemic infection. Current treatment protocols currently fail to treat nearly 1/3 of systemic candidiasis patients (34) and drug resistant strains of *C. albicans* have emerged in recent years (48). Thus, in order to develop new and effective anti-fungal drug targets, it is important to study the innate immune responses against *Candida albicans.*

v. Zebrafish as a Model Organism for the Immune System

Research concerning *C. albicans* that integrates immune system dynamics with comparative functional genomics is critical. Comparative genome studies indicate that there are conserved aspects of the innate immune system in all living systems, from *Drosophila* to human. And, due to practical considerations and ethical concerns over using mice in the laboratory, cost-effective model organisms such as the zebrafish (*Danio rerio*) are ever-increasing in popularity. Zebrafish embryos are transparent, allowing us to examine biological processes *in vivo.* Furthermore, zebrafish reproduce more quickly than mice (41) and share many conserved genes with humans (35, 41).

In particular, many components that play roles in immunity and development are conserved between zebrafish and humans. Complement components that play important roles in both arms of the immune system are highly conserved (20, 46). The pro-inflammatory cytokine interleukin-6 is also highly conserved between bony fish species

and humans (23). Essential development mechanisms in hematopoiesis including the notch (6) and hedgehog (50) pathways are highly conserved between fish and humans as well.

However, not all immune system components are fully conserved between fish and humans. Like mammals, zebrafish possess essential cells of the immune system, including macrophages, neutrophils, dendritic cells, NK cells, T-cells, and B-cells (43). Interestingly, it has been found that phagocytic B cells exist in all teleost fish species, but the demonstration of such B cells in mammals still remains elusive (43). Additionally, immunoglobulin light chain genes that contribute to antibody specificity are spread among five chromosomes in zebrafish instead of two, suggesting for more possible recombinations (56). Additionally, IgG, IgA, and IgE—three of the five major classes of anitbodies in mammals—are absent in fish, yet fish possess IgH and IgZ/T, both of which are absent in mammals (55). It has been demonstrated that teleost IgT is evolutionarily the earliest immunoglobulin protein and plays a pivotal role in gut mucosal immunity (43), suggesting IgT and mammalian IgA may possess functional similarities.

To complicate matters of conservation, genome studies have suggested that zebrafish genomes were once tetraploid instead of diploid (10, 35). Genomic mapping studies suggest large areas of conservation between different zebrafish chromosomes, so it is likely that at least one duplication event occurred early in evolutionary history (51). It has also been suggested that two polyploidization events occurred prior to this event (35). Additionally, some studies propose that three genome duplication events occurred prior to the emergence of zebrafish and that additional genome duplication events have governed much of evolutionary history (10). Comparative mapping studies have revealed

that alterations in gene order within linkage groups are common and that there is a lack of conserved gene order between the zebrafish and human genomes (51).

vi. Conserved Mechanisms of Innate Immune Recognition Between Zebrafish and Mammals

Pattern recognition receptors (PRRs) recognize pathogen-associated molecular patterns (PAMPs). PAMPs frequently consist of cell wall components (carbohydrate or protein) or genetic material (DNA or RNA), depending on whether the receptor is specific for viruses, bacteria, or eukaryotic organisms. Several examples of PRRs exist in vertebrates. The Toll-like receptors (TLRs) are a group of eleven transmembrane receptors that recognize nearly all types of pathogens. TLRs recognize numerous molecular patterns—from viral nucleic acids to lipids and carbohydrates (24). Activated TLRs induce the Toll-like receptor signaling pathway, which eventually leads to the transcription of pro-inflammatory cytokines through activation of MAP kinases or NF-κB (24). Scavenger receptors are another common PRR located on macrophages and dendritic cells. These receptors bind to and internalize bacteria and apoptotic host cells, triggering phagocytosis (24).

The C-type lectin receptors are PRRs that interact with a vast array of carbohydrates and possess a variety of functions (36, 51). There are currently seventeen recognized groups of C-type lectins, all of which vary in structure and function (54). Group II, III, IV, V, and VI C-type lectins play roles in immunity, and some Group II, III, V, and VI C-type lectins have been shown to recognize fungi in mammals. Group II C-type lectins are Type II transmembrane proteins characterized by a short cytoplasmic tail,

a transmembrane region, an extracellular stalk region, and a calcium- dependent carbohydrate recognition domain (CRD) (54). Group II C-type lectins are soluble collectins that possess an alpha-helix domain as well as a CRD (54). Group VI C-type lectins are Type I transmembrane proteins possessing N-terminal ricin-like and fibronectin-like domains, multiple extracellular CRDs, and a short cytoplasmic domain (54). Group V C-type lectins are similar to Group II C-type lectins in terms of basic structure; however, their CRD is not dependent on calcium ions (54). Although referred to as NK-like receptors (33), the vast majority of lectins in this subgroup are expressed on other immune cells such as macrophages, dendritic cells, and B-cells (54).

Several major fungal recognition C-type lectin receptors have been identified in mammals. Dectin-1 is a Group V C-type lectin receptor that is highly expressed on myeloid cells (36). It has been shown to recognize the inner beta-glucan layer of fungal cell walls (49). Dectin-1 possesses an immunoreceptor tyrosine activation-like motif (hemi-ITAM) in its cytoplasmic domain (36), which induces signal transduction upon recognition. Dimerized Dectin-1 has been shown to recognize beta-glucan through binding to the amino acid sequence WIH (9). The hemi-ITAM motif is then phosphorylated by spleen tyrosine kinase (Syk), inducing several signal transduction pathways leading to an innate immune response (32). Other Group V C-type lectins possess immunoreceptor tyrosine inhibition motifs (ITIMs), which block activation of signal transduction pathways and other functions (36).

Three prominent mammalian fungal receptors belong to the Group II subfamily of C-type lectins. Dectin-2, Mincle, and DC-SIGN (CD209) all recognize the outer mannan layer of fungal cell walls and are specific for mannoses. All are expressed on myeloid

cells and do not possess a functional ITAM or hemi-ITAM motif. However, Dectin-2 and Mincle have been shown to couple with the ITAM-containing FcRγ to induce signal transduction (38). Mannose-binding lectin (MBL) and macrophage mannose receptor (MR) are additional mannan receptors, yet MBL is a Group III C-type lectin and MR is a Group VI C-type lectin (54). Unlike Group V C-type lectins, Group II C-type lectins must interact with calcium ions in order to bind to their ligands (54), thus the presence of a calcium binding motif in these proteins is crucial to their function.

Although these fungal recognition receptors have been identified and characterized in mammals such as humans and mice, their presence in teleost fish species remains elusive. A homologue of MBL in the zebrafish has previously been identified (21). However, the presence of Group V C-type lectins such as Dectin-1 in teleost fish species has not been established. Comparative genome and phylogenetic studies of C-type lectins in the pufferfish, *Takifugu rubripes*, did not identify Group V C-type lectin receptors and found an overrepresentation of DC-SIGN-like proteins (53). Another study also failed to identify Group V C-type lectin receptors in zebrafish, but found Group II receptors possessing Group V properties that were phylogenetically related to Group C-type lectin receptors, and were classified as NK-like (33). Conversely, many Group II C-type lectins have been identified in bony fish, including DC-SIGN paralgoues in the pufferfish (53), as well as SCLRA, SCLRB, and SCLRC in Atlantic salmon (40). Despite this, the identification of Group II fungal recognition receptors remains elusive in fish. Although a potential DC-SIGN/CD209 homologue was identified in zebrafish (28), it has been poorly characterized. Homologues for Dectin-2 and Mincle in zebrafish also have

yet to be discovered. Despite the array of Group II C-type lectins identified in fish, the patterns that they recognize are currently unclear.

Herein, phylogenetic and syntenic analyses were employed to identify potential fungal recognition C-type lectin receptors in zebrafish and elucidate the evolutionary patterns of these immune receptors from fish to humans. Several C-type lectin pattern recognition receptors have been identified and cloned. Soluble fusion receptors for three attractive candidate receptors were generated in preparation for characterizing their specificity for recognizing fungi and other various microbes.

## MATERIALS AND METHODS

### i. Database Generation and Sequence Retrieval

A database containing known zebrafish C-type lectins was created using Microsoft Excel and the NCBI database. An iterative search was performed in order to obtain all known zebrafish genes containing a C-type lectin domain. "C-type lectin domain" was entered into the NCBI "Gene" search bar. The search results from all organisms were then narrowed down to just *Danio rerio* (zebrafish). From there, genes were selected one by one and added its information to the database, All Zebrafish CTLD Genes (All Zebrafish CTLD Genes, Electronic Appendix). Details added to the database included gene ID, gene name, NM/XM nucleotide identification numbers, NP/XP protein identification numbers, chromosome and location on chromosome, mRNA length, protein length, and presence of a hemi-ITAM (YXXL) sequence. This process was repeated for each of the zebrafish C-type lectin genes found in the NCBI database. An NCBI BLAST search against both the zebrafish and human genomes was subjected to all zebrafish

genes that were not manually annotated to ensure that they were potential C-type lectin receptors. Such genes whose BLAST hits included mammalian and zebrafish C-type lectins were added to the database.

ii. Determining Potential Group II and Group V C-type Lectin Receptors in Zebrafish

The data contained in the Microsoft Excel database were screened for probable Group II and Group V lectins. Genes and protein sequences whose names corresponded with lectins that did not belong to the Group II or Group V C-type lectin subfamilies were eliminated from the candidate list. The remaining protein sequences were analyzed for their likelihood of belonging to Group II or Group V based on BLAST hits and genome annotations. All lectin genes containing a DC-SIGN-like family annotation that were not collectins were added to the new database (Group II and V CTLD Zebrafish, Electronic Appendix). Candidate lectins were also aligned with mammalian C-type lectin receptors, salmon C-type lectin receptor, and zebrafish mannose-binding lectin using the COBALT multiple alignment tool (http://www.ncbi.nlm.nih.gov/tools/cobalt/).

iii. Membrane Topology Predictions

Candidate lectin transmembrane topologies were predicted using TMHMM Server 2.0 (http://www.cbs.dtu.dk/services/TMHMM/). The output format for TMHMM data was extensive with graphics (TMHMM, Electronic Appendix). TMPRED (http://www.ch.embnet.org/software/TMPRED_form.html) was also used to predict the presence and location of potential transmembrane-spanning domains. TMPRED parameters included a minimum transmembrane region length of 17 amino acids and a

maximum length of 33 amino acids. The output format for TMPRED data was in the form of html and PDF (TMPRED, Electronic Appendix). The data from these outputs were used to hypothesize the location of transmembrane regions and orientation of the N-terminus and C-terminus of all candidate fungal recognition receptors. Topology predictions from each algorithm were compared in order to hypothesize the topology and later construct the most accurate multiple alignment.

iv. Construction of Multiple Alignments

For each multiple alignment generated, sequences (Table 1) were aligned using ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Upon generating the alignment, each individual domain (cytoplasmic region, transmembrane region, extracellular neck region, and extracellular carbohydrate recognition domain) was aligned using T-Coffee (http://www.ebi.ac.uk/Tools/msa/tcoffee/). These alignments were then pieced together to generate one alignment that was manually edited in BioEdit. Sequences were then shaded according to 60% similarity. Relevant binding and signaling motifs, along with protein region labels, were outlined in Microsoft PowerPoint.

v. Construction of Phylogenetic Trees

All phylogenies were constructed using the carbohydrate recognition domain (CRD) amino acid sequences from human, murine, and chicken Group II and Group V C-type lectin receptors and potential Group II or Group V receptors in zebrafish, puffer, and salmon (Table 1). All sequences were initially retrieved from the NCBI database and edited into CRD sequences using Notepad and BioEdit. The desired CRD amino acid

sequences were then aligned using the ClustalW component of the MEGA5 package (44). Output alignment data were used to construct neighbor-joining phylogenies. Neighbor-joining trees were generated in MEGA5 using the Poisson distribution model and 2000 bootstrap replications. Tree topologies were organized and edited using M5 Tree Explorer (44). All trees were generated in triplicate to ensure repetition of data.

vi. Synteny Analysis

The genomic location data for zSCLRA, zKCLectin, Dectin-1, Dectin-2, Mincle, CD209, DC-SIGNR, and zDC-SIGN-like F were retrieved from the NCBI and Ensembl online databases. All genomic location data was obtained from the NCBI genomic regions, transcripts, and products data and Ensembl location data. Human genome data was taken from the GRCh37.p10 primary genome assembly. Mouse genome data was taken from the GRCm38.p1 primary genome assembly. Zebrafish genome data was taken from the Zv9 primary genome assembly. Genomic location data was used to construct images depicting the genomic location of these genes. These images were produced in Microsoft Power Point and Adobe Illustrator. Percent homology between the syntenic genes was determined through NCBI BLAST. For the purposes of synteny analysis, amino acid homology values exceeding 50% were considered significant.

vii. Establishing a Zebrafish Head Kidney cDNA Library

Ten adult AB zebrafish were sacrificed in accordance to the IACUC protocol A2012-11-03. Head kidneys were dissected from the euthanized zebrafish and homogenized in TriZol reagent. RNA was extracted from homogenized head kidneys

using the RNeasy RNA extraction kit (Qiagen). RNA was quantified and assesses for purity using a nanodropper. A cDNA library was synthesized from the extracted RNA according to the ImProm-II Reverse Transcription Protocol.

viii. Obtaining PCR Amplicons

The carbohydrate recognition domain (CRD) regions of zSCLRA, zDC-SIGN-like F, and zKCLectin were amplified using the polymerase chain reaction (PCR) with the CRD-specific primers listed in Table 2. Phire polymerase (NEB) was utilized for the reaction. GAPDH forward and reverse primers were added to two cDNA samples as positive controls. GAPDH primers were also incubated in reaction mixes containing sterile nuclease-free water instead of cDNA as negative controls. cDNA samples were denatured at 98°C for 15 seconds, annealed at 58°C for 15 seconds, and elongated at 72°C for 40 seconds. 40 cycles of PCR were performed. All PCR amplicons were stored at -20°C. Amplicons were detected through gel electrophoresis by loading samples onto a 2% agarose gel in 1X TAE. Amplicons were purified from the gel according to the QIAquick gel extraction kit protocol (Qiagen).

ix. Cloning Potential Fungal Recognition Receptor Amplicons

Purified PCR amplicons were cloned into the pGEM-T easy vector (Promega) and transformed into *E. coli* NEB5α chemically competent cells according to the manufacturer's instructions. Positive FRR-pGEM-T transformants were selected and grown in Terrific Broth with Ampicillin (100 μg/ml) at 37°C overnight. FRR-pGEM-T plasmids were purified using the QIAquick mini-prep kit (Qiagen). Purified plasmids

were sequenced at the University of Maine DNA Sequencing Facility to detect plasmids containing FRR insert sequences matching the genomic DNA sequences provided on the NCBI database (FRR-pGEM-T Clone Sequences, Electronic Appendix).

Plasmids whose DNA matched that of database sequences and Dectin-2-pSecTag2 (30) were digested with KpnI and EcoRI. Digests were loaded onto a 1% agarose gel. The pSecTag2 and FRR amplicon bands were extracted using the QIAquick gel extraction kit (Qiagen). The FRR CRD inserts and pSecTag2 were ligated together using T4 DNA ligase and T4 DNA ligase buffer (NEB). *E. coli* NEB5α chemically competent cells were transformed with FRR-pSecTag2 plasmids according to the manufacturer's protocol. Transformants were selected and grown in Terrific Broth with Ampicillin (100 μg/m) overnight at 37°C. FRR-pGEM-T plasmids were purified using the QIAquick mini-prep kit (Qiagen). Plasmids containing a single insert were identified by digesting purified plasmids with EcoRI. Plasmids containing a single EcoRI restriction enzyme site were detected by digesting purified plasmids with SpeI. Plasmids were then sequenced at the University of Maine DNA Sequencing Facility to detect plasmids containing FRR insert sequences matching the genomic DNA sequences provided on the NCBI database (FRR-SecTag2 Clone Sequences, Electronic Appendix)

*E. coli* NEB5α containing the FRR-pGEM-T and the FRR-pSecTag2 plasmids were cryopreserved at -80°C in 25% glycerol. Cells containing the FRR-pSecTag2 plasmids were grown on L-agar Ampicillin (100 ug/mL) plates and incubated overnight at 37°C. Colonies from the plate were subcultured in Terrific Broth with Ampicillin (100 μg/ml) overnight at 37°C. Broth cultures were subjected to endonuclease free mini-preps (MO BIO) according to the manufacturer's protocol.

x. Nucleofection and Cell Culture

HEK293T cells were grown in DMEM + Glutamax media containing 10% FBS, 1% Penicillin/Streptomycin, 1% sodium pyruvate, and 1% gentamycin. Cells were maintained in the presence of $CO_2$ at 37°C and split when flasks reached 80-90% confluency. Cells ($1 \times 10^6$) were nucleofected according to the Amaxa Cell Line Nucleofector Kit V protocol (Lonza) and using a Nucleofector II Device. Cells were nucleofected with 5 uL of zSCLRA DNA in 100 uL of Cell Line Nucleofector Solution V (Lonza). Nucleofected cells were selected for in zeocin (100 mg/ml) and later transferred to a CELLine Bioreactor membrane culture flask (CELLine) to isolate secreted Fc fusion protein.

xi. Purification of Soluble Fc Fusion Proteins

Cell culture supernatants containing secreted protein were applied to HiTrap Columns containing Protein A (Sigma-Aldrich Co.) for FRR-Fc purification. Soluble protein was purified according to the manufacturer's protocol. Eluted protein was dialyzed in a Slide-A-Lyzer 20,000 MWCO cassette (Pierce). Dialysis was performed overnight at 4°C in 1X PBS with stirring. Protein samples were concentrated using Amicon Ultra 0.5 ml 3k microconcentrators (Millipore) according to manufacturer's instructions. Purified and concentrated protein was stored at -20°C in 45% glycerol. Protein was quantified with a Bradford Assay using a Bovine Serum Albumin (BSA) standard curve.

xii. Immunofluorescence Staining

Wild-type GFP (WT-GFP) SC5314 (47) and KAH3-GFP (18) *Candida albicans* strains were inoculated onto YPD plates (2% Bacto Peptone, 1% Bacto Yeast Extract, 2% Glucose, 2% Bacto Agar). WT-GFP *C. albicans* were grown overnight at 30°C, while KAH3-GFP *C. albicans* were grown for 36-48 hours at 30°C. One day prior to immunofluorescence staining, both strains of *C. albicans* were grown overnight in 4 ml YPD liquid (2% Bacto Peptone, 1% Bacto Yeast Extract, 2% Glucose). Stocks of both *C. albicans* strains were cryopreserved at -80°C in 50% glycerol and YPD liquid.

Both WT-GFP (300 µl/sample) and KAH3-GFP (600 µl/sample) were centrifuged at 21000 x g for 30 seconds. All cells were resuspended and washed three times in 1X PBS with 1 mM $CaCl_2$. Cells were then blocked with 2% BSA in 1X PBS with 1 mM $CaCl_2$ at room temperature for 30 minutes. Next, samples were stained with 3.3 ng/µl Dectin-1-Fc (42), 8.25 ng/µl zSCLRA-Fc, 3.3 ng/µl Dectin-1-Fc plus 100 µg/ml laminarin (Sigma-Aldrich), 8.25 ng/µl zSCLRA-Fc plus 100 µg/ml laminarin, or human IgG Fc control protein (Millipore) at 4°C for one hour. Cells were resuspended and washed three times in 1X PBS with 1 mM $CaCl_2$. Samples were then stained with 1.33 ng/µl Cy3 DαH (donkey anti-human) secondary antibody (Jackson ImmunoResearch) at room temperature for 20 minutes. Cells were resuspended and washed three times in 1X PBS with 1 mM $CaCl_2$ before visualized under a fluorescence microscope.

## RESULTS

### i. Initial Construction and Screening of a Zebrafish C-type Lectin Database Revealed Three Potential Fungal Recognition Receptors

To establish a pool of all known zebrafish C-type lectins, an Excel database containing zebrafish C-type lectins was generated using data from NCBI. Search criteria included "C-type lectin domain" and were restricted to *Danio rerio.* BLAST hits revealed more C-type lectins in the zebrafish. The data were then screened for similarity to Group II and Group V C-type lectins based on BLAST hits and basic alignments, prioritizing potential fungal recognition receptors based on the presence of specific mannose or beta-glucan recognition motifs, signaling motifs, and genome annotations.

Genes whose names corresponded with non-group II and V C-type lectins were eliminated from the list, as these are not related to Dectin-1 or Dectin-2. This reduced the list to fewer candidate genes, including two novel zebrafish proteins potentially similar to NKRs. Upon finding the final candidate proteins, genes were aligned against at least one member from other zebrafish CTLDs and also against SCLRA, murine Dectin-1, and human Dectin-1 using the NCBI BLAST COBALT alignment tool to confirm similarity. After this, three candidate lectins with a high potential of binding fungi remained. These included zebrafish SCLRA (zSCLRA), zebrafish killer cell lectin 1B (zKCLectin), and zgc:174904. Zgc:174904 was among multiple DC-SIGN-like lectins and will be referred to as zDC-SIGN-like F (DCSF).

ii. The Carbohydrate Recognition Domain (CRD) of Zebrafish DC-SIGN-like F Shares Homology With Zebrafish CD209, Human CD209, and Human DC-SIGNR

To analyze the similarity of zDC-SIGN-like F to other C-type lectins in the DC-SIGN family, zDC-SIGN-like F was aligned with zebrafish CD209, human CD209, and human DC-SIGNR using ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/). BioEdit was then used to manually modify the alignment so that it matched the TMPRED and TMHMM topology data (Figure 1). The alignment revealed that the cytoplasmic region of zDC-SIGN-like F is similar in length to the cytoplasmic region of both human CD209 and DC-SIGNR. The cytoplasmic region of zCD209 is much longer than those of the other proteins in the alignment. The tri-acid cluster present in human CD209 and DC-SIGNR was found in both zCD209 and zDC-SIGN-like F. The leucine repeat present in human CD209 and DC-SIGNR was absent in zDC-SIGN-like F. A full list of conserved residues and motifs can be found in Table 3.

The transmembrane regions of these lectins were variable in amino acid composition, although some patterns were present. All candidate lectins lacked transmembrane or cytoplasmic domain arginine residues, suggesting that the lectins would not couple with the FcRγ to induce signal transduction. This mechanism occurs in Dectin-2 (19). Furthermore, the transmembrane regions were similar in size and predominantly hydrophobic, as confirmed by the TMPRED hydrophobicity plots (Figures 2 and 3). Likewise, there was little conservation present in the extracellular neck region of these lectins. The neck regions of human CD209 and DC-SIGNR were significantly longer in length than those of zDC-SIGN-like F and zCD209 and contained more repeats.

The extracellular carbohydrate recognition domain (CRD) of zDC-SIGN-like F shares much homology with the CRD of the other lectins. The WxGL (WIGL) carbohydrate recognition sequence was partially conserved in the extracellular CTLD, as the L was missing in zDC-SIGN-like F and the full WIGL was present in zCD209 only. An EPN mannose recognition sequence (40) was conserved in the CRD between zDC-SIGN-like F, zCD209, and human CD209 and DC-SIGNR. The WND carbohydrate recognition sequence and calcium binding site (54) was conserved in the extracellular domain among the mammalian DC-SIGN proteins, but this domain wasn't completely conserved in the zebrafish proteins. WTD and WYD were present in zCD209 and zDC-SIGN-like F respectively. A full list of conserved residues can be found in Table 3.

iii. The Mammalian CRD Region and Immunoreceptor Tyrosine Activation Motifs (ITAMs) are Conserved in zKCLectin and zSCLRA

The predicted zKCLectin and zSCLRA polypeptides are shorter than zDC-SIGN-like F and more similar to the lengths of the Dectin-1 and Dectin-2 proteins. Thus, zKCLectin and zSCLRA were aligned against salmon SCLRA, human Dectin-1, Dectin-2, and Mincle to determine sequence homology using ClustalW. BioEdit was then used to manually modify the alignment so that it matched the TMPRED and TMHMM topology data (Figure 4). A hemi-ITAM or ITAM-like sequence (36) is present in the cytoplasmic region of both human and murine Dectin-1 and contributes to Syk signaling (14). The alignment revealed that potential ITAM-like sequences lie in the cytoplasmic regions of zKCLectin, zSCLRA, and sSCLRA. TMPRED and TMHMM data confirmed that the predicted cytoplasmic region of zKCLectin is significantly longer than the others at 82

amino acids. Dectin-2 and Mincle have cytoplasmic regions of 20-22 amino acids, whereas the other lectins in the alignment have cytoplasmic regions between 40 and 50 amino acids in length.

The transmembrane regions of these proteins were heavy in hydrophobic amino acids and were approximately 20 amino acids in length (Figure 4). Mincle contained a transmembrane arginine residue, which was absent in the other lectins in this alignment. The neck regions of these lectins were variable in both length and amino acid composition. There was conservation in the CRD regions, however. All extracellular CRD regions began with a cysteine residue. The WIGL carbohydrate recognition sequence (54) was conserved in each of the lectins except for Dectin-2 and Mincle. The EPN mannose recognition motif (40) was present in each lectin except for Dectin-1, zKCLectin, and zSCLRA. Instead, the Dectin-1 lectin contained a WIH beta-glucan recognition sequence (9). The WND calcium binding sequence (54) was present in Dectin-2 and Mincle only. A WFD sequence was substituted for both sSCLRA and zSCLRA, whereas this sequence was missing in both Dectin-1 and zKCLectin altogether. Furthermore, the predicted CRD region for zKCLectin was shorter than that for the other lectins in this alignment.

iv. Synteny Patterns Suggest Common Ancestry for Some Zebrafish C-type lectins and Mammalian Immune C-type Lectins

A synteny analysis was performed to determine the level of genomic conservation between the NK cluster on human chromosome 12, which includes Dectin-1, Dectin-2, and Mincle, and the locations of the potential fungal recognition receptors whitin the

zebrafish genome. Zebrafish chromosome 16 encodes zSCLRA and zKCLectin, whereas zDC-SIGN-like F is located on chromosome 19. The synteny analysis was performed by comparing the genomic location of the potential fungal recognition receptors to that of their potential mammalian homologues. It was found that zebrafish chromosome 16 also encodes the zebrafish homologues to Pex5, Clstn3, Lpcat3, Gnb3, Cops7a, and ZNF384. The human homologues of these genes are encoded for on human chromosome 12 and lie just 1 Mb upstream of the NK cluster where Dectin-1, Dectin-2, and Mincle are located (Figure 5). These homologues are encoded for on mouse chromosome 6 in the middle of the NK-cluster. This demonstrates that while the overall gene locations are conserved, gene order isn't necessarily conserved.

A percent homology analysis was performed on all syntenic genes to illustrate the relatedness of each homologue and potential homologue. Syntenic genes not associated with fungal recognition possessed highly conserved amino acid sequences in zebrafish, humans, and mice (Tables 4a-4c). Overall, zSCLRA shared the most sequence homology with human Dectin-1 (Table 4a). Both zKCLectin and zDC-SIGN-like F shared the most sequence homology with human CD209 (Table 4a). Human and mouse proteins shared the most homology in amino acid composition (Table 4c).

Synteny patterns between zDC-SIGN-like F and other potential fungal recognition receptors were not detectable. There is a cluster of immune-related lectin-like receptors (Illrs) located on zebrafish chromosome 19 (33) that could be homologous to the NK cluster on human chromosome 12. However, no syntenic relationships between the genes flanking the Illr cluster and the genes flanking the NK cluster were detected, so these relationships are currently inconclusive. Furthermore, zDC-SIGN-like F is also

encoded on zebrafish chromosome 19. Additional DC-SIGN-like proteins in zebrafish possessed no synteny with the NK cluster on human chromosome 12 or with other Group II or Group V C-type lectin patterns in mammals (not shown). We attempted to enlarge our synteny searches to the chicken genome. However, not enough information regarding genomic location of Group II and Group V C-type lectin receptors in chicken was available for analysis.

v. NK-like Receptors Evolved Independently in Teleost Fish From Mammalian and Teleost Group II C-type Lectin Receptors

Two phylogenetic trees were constructed using MEGA 5.0 in order to characterize the evolutionary relationships between potential fish fungal recognition receptors and identified mammalian fungal recognition receptors. One phylogeny consisted of the CRD regions of potential teleost Group II and Group V lectins aligned with human, murine, and chicken Group II and Group V lectins (Figure 6). The second phylogeny consisted of the CRD regions of potential Group II and Group V teleost C-type lectins (Figure 7).

The first phylogeny revealed that mammalian Group V C-type lectin receptors and teleost NK-like receptors evolved independently from Group II C-type lectin receptors. It can be inferred from this tree that there appear to be no definitive Group V C-type lectin receptors in teleost fish species. The closest teleost ancestors to the Group V C-type lectins are the NK-like receptors. These include sSCLRA, zSCLRA, and the Illrs. The NK-like receptors are bracketed next to the mammalian and chicken Group V C-type lectins, which suggests that either both families evolved independently from

Group II C-type lectins or that NK-like receptors are a hybrid of Group II and Group V C-type lectins. The lectins in this phylogeny clustered together based on both function and species. Human and murine Dectin-1 clustered together, along with human and murine CD209. The non-mammalian chicken outgroups of Group V lectins also clustered together. The Dectin-2 family (Dectin-2, Mincle, DC-SIGN) clustered together with the Group II C-type lectins. Furthermore, SCLRC and zDC-SIGN-like F were clustered together within the Dectin-2 family.

The second phylogeny confirmed that NK-like receptors evolved independently of Group II C-type lectins in fish. High bootstrap values were obtained for most of the relationships and the NK-like receptors described above clustered together. The majority of the phylogenetic relationships were based on the function of the protein as opposed to the species, as zebrafish and salmon SCLRA clustered together. Salmon SCLRC and zebrafish zDC-SIGN-like F also clustered together independently of zSCLRA and sSCLRA, suggesting homology between SCLRC and zDC-SIGN-like F. Fugu DC-SIGN paralogues were dispersed throughout the phylogeny. Based on both phylogenies, it is evident that zDC-SIGN-like F and zKCLectin appear to be more similar to Group II C-type lectins than NK-like C-type lectins based on phylogenetic relationships.

### vi. FRR-pSecTag2 Vectors Code for Expression of a Soluble Fungal Recognition Receptor Protein

In order to generate the soluble fusion proteins zSCLRA-Fc, zKCLectin-Fc, and zDC-SIGN-like F-Fc, the carbohydrate recognition domain (CRD) of each candidate receptor was cloned into pSecTag2 (Figure 8). This plasmid fuses the CRD to both a

secretion signal sequence and the constant portion of human Fc. The signal sequence ensures the protein will be secreted into the media and the Fc is designed to both enhance solubility and permit easy detection using inexpensive fluorescently-labeled anti-Fc antibodies.

Because zebrafish head kidneys, analogous to the bone marrow, contain the highest concentration of myeloid cells, RNA was extracted from 10 homogenized zebrafish head kidneys and a cDNA library was synthesized. The CRD regions of zSCLRA, zKCLectin, and zDC-SIGN-like F were amplified using PCR and CRD-specific primers (Table 2). Amplicons were purified and cloned into pGEM-T in order to check the PCR product sequences against the predicted cDNA sequences given on the NCBI database (FRR-pGEM-T Clone Sequences, Electronic Appendix). After checking sequences, the FRR CRDs were then subcloned into the pSecTag2 vector (30) by replacing the Dectin-2 insert with the FRR insert to obtain zSCLRA-pSecTag2, zDC-SIGN-like F-pSecTag2, and zKCLectin-pSecTag2. An Igκ secretion signal sequences lies upstream of the insert in the pSecTag2, whereas a sequence encoding Fc lies downstream of the insert (30). When the FRRs are nucleofected and overexpressed in HEK293T cells, a soluble FRR-Fc fusion protein should be secreted into the medium.

vii. Purified zSCLRA-Fc is Concentrated Enough for Ligand Binding Assays

In order to generate protein usable for future *C. albicans* binding assays, secreted zSCLRA-Fc generated in the CELLine Bioreactor membrane culture flask was purified over protein A agarose. Protein A is produced by *Staphylococcus aureus* and binds to the Fc region of IgG (39). The purified protein was then quantified using a Bradford Assay.

Quantification data revealed that the concentration of zSCLRA-Fc isolated was 0.5208 (μg/μl). Given that a final concentration of 3.3 ng/μl is the desired concentration of Dectin-1-Fc in immunofluorescence staining assays, the purified zSCLRA-Fc was concentrated enough to be utilized in *C. albicans* binding assays and carbohydrate competition assays.

## DISCUSSION

### i. Overview

The innate immune system is highly conserved amongst all multicellular organisms, and evolves rapidly to adapt to the constant battle between host cells and pathogens. I used *in silico* methods to elucidate the evolutionary patterns of C-type lectins to understand better how zebrafish respond to pathogenic fungi such as *Candida albicans*. Mammals such as humans and mice possess conserved C-type lectin receptors that recognize the *C. albicans* cell wall. Based on homology to mammalian and fish C-type lectins *in silico* methods were employed to identify three potential fungal recognition receptors in zebrafish. Multiple alignments revealed that several important sequence motifs involved with fungal carbohydrate recognition and signal transduction are conserved between the potential teleost fungal recognition receptors and the known mammalian receptors. This suggests that the functions between these receptors may be conserved. Syntenic relationships also exist between zSCLRA, zKCLectin, and human Dectin-1, Dectin-2, and Mincle, suggesting that the functions of these receptors may be conserved in zSCLRA and zKCLectin. Phylogenetic analysis determined that all of the

candidate zebrafish C-type lectin fungal recognition receptors (FRRs) are Group II C-type lectins, confirming previous studies that were unable to identify Group V C-type lectin receptors in teleosts. *In vitro* assays using soluble FRR-Fc proteins are currently being employed to analyze the fungal recognition properties of these receptors. In the event that these experiments fail to identify an FRR-Fc receptor for mannan or beta-glucan, alternative *in vitro* or *in vivo* assays can be performed to further analyze potential fungal recognition.

ii. Sequence Homology is Present Amongst Mammalian and Potential Teleost Fungal Recognition Receptors

Several elements of innate immune C-type lectin receptors are conserved from fish to mammals. A potential hemi-immunoreceptor tyrosine activation motif (hemi-ITAM) was detected in zSCLRA, sSCLRA, and zKCLectin, but was absent among other potential zebrafish fungal recognition receptors. Both SCLRA homologues contained a YTSL sequence, and zKCLectin contained a YSSL sequence. An ITAM is present in Dectin-1 and is critical for Syk-mediated signaling (14). A YXXL sequence doesn't guarantee the presence of a functional ITAM, as a YSKL sequence is present in human DC-SIGNR and CD209, but it doesn't contribute to Syk signaling (citation). An ITAM is absent in Dectin-2 (5) and in Mincle; however, signaling occurs with coupling of the FcRγ (38) to a transmembrane or cytoplasmic arginine residue (16).

The extracellular carbohydrate recognition domain (CRD) is highly conserved between fish and mammals. A WIH sequence that specifically contributes to beta-glucan recognition (4, 9) was absent among the potential fungal recognition receptors, although a

WIH sequence in the beginning of the CRD was found in sSCLRA. The EPN sequence is specific for mannose recognition (40, 51). This sequence was present in zDC-SIGN-like F, zSCLRA, zCD209, and several other potential fungal recognition receptors. This is consistent with the presence of an EPN sequence in human Mincle, Dectin-2, CD209, and DC-SIGNR. These receptors have been shown to bind to mannoses and recognize the mannoprotein layer present in the fungal cell wall. Furthermore, the EPN motif may contribute to calcium binding.

Interestingly, an EPD sequence is present in sSCLRA (40) and in the Illrs (33, 52). Multiple alignments suggest that an aspartate residue is substituted for an asparagine residue, as EPD is present in the same place as EPN in the sequence. However, glutamate may be a substitution for glutamine in a QPD galactose recognition sequence (40). Furthermore, EPD may also contribute to calcium binding in teleost Group II C-type lectins (40). Carbohydrate competition assays performed with fungi and bacteria would be necessary to determine the exact function of the EPD sequence.

Overall, the multiple alignments reveal that several critical sequence motifs such as hemi-ITAMS and EPN sequences are conserved in some of the potential zebrafish receptors. Despite this, it is possible that these motifs are not involved in pattern recognition or are non-functional. Therefore, these potential receptors must be characterized *in vivo* or *in vitro* to determine their exact fungal recognition abilities.

iii. Synteny Patterns Are Sometimes Present Between Fish and Mammalian C-type Lectin Receptors

Synteny patterns were not always identified between potential fish FRRs and human FRRs. A cluster of Illr genes appears on zebrafish chromosome 19 (33). However, homologues of genes in the NK cluster surrounding the Illr cluster are absent. Predicted zebrafish CD209 appeared on a separate chromosome from other potential FRRs and there were no synteny patterns present between zebrafish CD209 and human CD209. Additionally, zDC-SIGN-like F is encoded by zebrafish chromosome 19. The genes surrounding this have human homologues on human chromosome 17 that are highly conserved and surround CLEC10A. This may suggest homology between CLEC10A and zDC-SIGN-like F, even though the CRD of zDC-SIGN-like F is highly homologous to that of Mincle and Dectin-2.

zKCLectin and zSCLRA share the highest conservation of genomic location with mammalian fungal recognition receptors. Both are encoded by zebrafish chromosome 16 and are flanked by several genes whose human homologues appear about one megabase upstream from the NK cluster. Murine homologues of these genes are encoded for in the middle of the NK cluster on mouse chromosome 6, suggesting that the NK cluster is highly conserved, even though the order of the genes is not always conserved in the NK-cluster.

The percent sequence homology results obtained from BLAST hits confirmed that the majority of the syntenic genes between zebrafish, mice, and humans can be considered homologues. This is particularly true for non-CTLD genes. However, zSCLRA shared the highest percent homology with human Dectin-1 and zKCLectin

shared the highest percent homology with human CD209. This contradicted the sequence homology results obtained from the multiple alignments, as one would expect zSCLRA to be more homologous with Dectin-2 or Mincle based on its conserved motifs and one would expect zKCLectin to be more homologous with Dectin-1 due to the absence of an EPN motif and presence of a hemi-ITAM. These disparate results could result from the fact that NCBI BLAST does not take into account conserved motifs or the location of these motifs with respect to the cell membrane.

In general, there is high genomic conservation between fish and mammals, but exact synteny patterns between fish and mammals are unknown. It has been demonstrated that C1qa, C1qb, and C1qc is highly conserved from fish to mouse to human (20). However, there is no sign of synteny in this region, suggesting that there may not be syntenic relationships between zebrafish and human complement genes. Furthermore, the presence of an NK cluster in fish remains elusive. There is an Illr cluster in zebrafish (33); however, there is little or no synteny present surrounding this cluster, suggesting that it may not be analogous to the NK cluster. Unfortunately, synteny patterns among zebrafish, chicken and platypus were undetectable, as the exact genomic locations of many C-type lectin receptor genes in chicken and platypus is unknown (1, 2).

Synteny patterns weren't consistently identified between the potential zebrafish fungal recognition receptors and mammalian fungal recognition receptors. Although some synteny existed between the human NK cluster on Chromosome 12 and zSCLRA and zKCLectin, the majority of potential zebrafish fungal recognition receptors do not share synteny with Dectin-1, Dectin-2, Mincle, or DC-SIGN. While the syntenic relationships suggest that homologues of such receptors may be absent in zebrafish

altogether, the lack of synteny may be due to the large evolutionary distances between zebrafish and mammals.

iv. Evolutionary Patterns of Potential Fungal Recognition Receptors are Consistent with Previous Phylogenies

In constructing phylogenetic trees, it is important to identify appropriate outgroups to bridge the evolutionary distances between teleost fish and mammals (3). In testing outgroups, it was found that chicken and platypus were the most suitable for the phylogeny constructed, especially when the extracellular C-type lectin domains were aligned prior to phylogeny generation. While frog is a common outgroup among phylogenies, few Group II and Group V C-type lectins have been identified in frog. When the few lectins detected were included in the phylogeny, a tree with low bootstrap values was generated (not shown), indicating that frog is a poor outgroup for determining the evolutionary patterns between fish and mammalian innate immune C-type lectins.

When Group II and Group V chicken or platypus lectins are used in combination with human, mouse, and teleost innate immune C-type lectin receptors, there appeared to be no Group V C-type lectins in fish, although the predicted NK-like Group II lectins clustered near the Group V C-type lectins. Platypus was ultimately eliminated from the original phylogeny. Adding ten additional zebrafish DC-SIGN-like proteins to the phylogenetic alignment caused the alignment file to be too large for MEGA to process, but including chicken only still allowed MEGA to process the file. It is hoped that future alignment tools will permit more detailed analyses.

The evolutionary patterns found in the phylogenies are consistent with previous experiments conducted. It was confirmed that pufferfish do not possess Group V C-type lectins (53) and that DC-SIGN-like proteins in pufferfish are overrepresented (53). The salmon C-type lectins A, B, and C (SCLRA, SCLRB, and SCLRC) were found to be Group II C-type lectins, although SCLRA and SCLRB were found to be NK-like Group II lectins. The zebrafish Illrs (33) and DC-SIGN (28) are also Group II according to the phylogeny. Thus, no Group V C-type lectins in zebrafish have been identified to date. There were also an abundance of DC-SIGN-like proteins in zebrafish that haven't been fully characterized, suggesting that, like in pufferfish, DC-SIGN homologs may be overrepresented in zebrafish.

A large number of Group II C-type lectins appeared to be closely related to Group V C-type lectins. These have been previously described as Group II C-type lectins with Group V properties, or NK-like (33). This suggests that there may be a hybrid C-type lectin group unique to teleost fish that possesses characteristics of both Group II and Group V C-type lectins, but doesn't fit into either category exactly. This small bracket of lectins appears to be consistent across all phylogenies tested, as well as in the previous literature.

Other evolutionary patterns evident between fish and humans mirror some of the patterns observed for the C-type lectin receptors. IL-6 and gp130 activation of STAT3 was found to be highly conserved between mammals and teleosts (23). Furthermore, complement components (Cq1a, Cq1b, and Cq1c) are conserved from fish to humans and are present in other species such as chicken, dog, and rat (20). IgA, IgE, and IgG are all missing in fish spcies; however, an IgZ/T exists in fish that is absent in mammals (55).

The absence of these immunoglobulins in fish could potentially mirror the absence of Group V lectins in fish and the presence of the anti-freeze lectins that are unique to fish (54). Furthermore, there may be a hybrid NK-like C-type lectin receptor in fish that possesses characteristics of Group II and Group V lectins.

Interestingly, the results from sequence homologies and phylogenies suggest that not only are there no Group V lectins in zebrafish, but that there is no apparent Dectin-1 homolog in zebrafish. Studies have shown, however, that the high-hyphal mannose receptor Dectin-2 plays a prominent role in murine fungal infections. Therefore, it is possible that a beta-glucan receptor is unnecessary for fungal recognition in the zebrafish, especially if there is an over-representation of potential mannose receptors. Furthermore, if Group V C-type lectin receptors did not evolve in zebrafish, it is possible that there is a beta-glucan receptor that is not a Group V C-type lectin. To test for this, one could inject zebrafish with beta-glucan extracts and monitor an immune response to the antigen. If such a receptor is present, the zebrafish would mount an immune response and clear the beta-glucan from its system.

Our phylogenetic analyses confirm the absence of Group V C-type lectins and the presence of NK-like receptors in teleost fish species. Thus, it is possible that a hybrid C-type lectin that shares characteristics of both Group II and Group V C-type lectins evolved in zebrafish. The phylogenies also suggest a possible over-representation of DC-SIGN-like proteins in zebrafish, which may be due to gene duplication events. While the phylogenies suggest that most zebrafish Group II C-type lectins are evolutionarily distant from mammalian fungal recognition receptors, it is still possible that a structurally different mannan or beta-glucan receptor exists in zebrafish.

v. Current and Future Work to Assay the Roles of Predicted Immune Receptors

*In silico* analysis alone is not enough to determine the specific pattern recognition properties of potential zebrafish FRRs. Thus, one must employ *in vitro* or *in vivo* assays to determine the specific properties of these potential receptors.

We cloned and established stable cell lines expressing zSCLRA-Fc. A soluble zSCLRA-Fc chimera receptor was successfully isolated from the cell culture medium and purified. Soluble zKCLectin-Fc and zDC-SIGN-like F-Fc are currently being developed in the same manner as zSCLRA-Fc. After purification, the specific binding properties of the potential FRRs can be examined by employing carbohydrate competition assays and microbial binding assays.

The purified receptors can be screened for their pattern recognition abilities. An assay that examines the binding of fluorescently-tagged *C. albicans* or zymosan to the receptors can be employed according to the immunofluorescence protocol previously listed. The binding properties of zSCLRA-Fc are currently being analyzed using immunofluorescence staining. Preliminary data suggests that zSCLRA-Fc does not recognize fungal mannans or beta-glucans. Despite this, it is possible that zSCLRA-Fc may recognize gram-positive or gram-negative bacteria or self-ligands. This receptor, along with the others, can also be tested for binding to gram-positive bacteria and gram-negative bacteria. Ligand binding can be observed qualitatively through immunofluorescence microscopy or quantitatively through flow cytometry. These results can be used to determine whether the CRDs of the potential fungal recognition receptors recognize fungi or bacteria.

In the future, it would be ideal to perform a Western Blot after purification of the FRR-fusion proteins. A Western Blot compares the protein content of the purified protein, unpurified supernatant, and intermediate fractions collected during purification in order to assess the purity of the isolated protein. This step was omitted from the purification procedure due to time constraints.

Competition assays can also be performed to determine ligand specificity. Receptors can be incubated with latex beads containing varying concentrations of fluorescently-labeled fungal and bacterial carbohydrates such as galactose, glucose, and mannose. Binding can be observed through immunofluorescence microscopy or flow cytometry. This assay can determine not only the specificity of the carbohydrates, but also the concentrations at which ligand binding is optimal or inhibited.

It is also possible that the Fc domain of the fusion proteins interferes with the binding specificity of the candidate FRRs. If this were the case, the full FRR genes could be cloned into pGEM-T as previously described. The plasmid could then be transfected into HEK293T cells. These cells do not normally recognize fungal sugars, and we will test for binding to fungi by looking for binding of fluorescent labeled *C. albicans* cells. In the event that zSCLRA, zKCLectin, and zDC-SIGN-like F did not recognize fungi, it would be ideal to revisit the zebrafish Group II C-type lecitn database (Group II and V CTLD Zebrafish, Electronic Appendix) and identify other lectins that may potentially recognize fungi. These lectins could then be cloned and expressed in the same way as previously described.

The role of potential FRRs in zebrafish candidiasis can also be examined *in vivo.* Morpholinos can be designed to knock-down expression of the candidate receptors.

Morpholinos are gene-specific anti-sense oligonucleotides that block gene expression in zebrafish embryos and larvae. Upon injection of the morpholinos into the zebrafish embryos, embryos can be infected with *C. albicans* and mortality curves can be constructed for FRR-knockdown embryos and wild-type embryos. Higher mortality curves in the absence of the potential FRR would indicate that the candidate gene is necessary for immunity to *C. albicans*. Morpholinos could also be designed that knock-down expression of intermediates in the signal transduction pathways initiated by FRRs that bind to *C. albicans* to determine the role of these pathways in zebrafish fungal immunity.

Overall, there are multiple *in vitro* and *in vivo* approaches to characterizing the biological roles of zSCLRA, zKCLectin, and zDC-SIGN-like F. While these receptors may not play roles in fungal recognition, it is possible that they recognize bacteria or self-ligands. Zebrafish possess a number of other Group II C-type lectin receptors that have not been characterized. In the future, it may be useful to revisit some of these receptors and perform more extensive analysis of their potential fungal recognition properties.

vi. Conclusion

Our *in silico* results reveal that three zebrafish C-type lectins—zSCLRA, zKCLectin, and zDC-SIGN-like F—are attractive candidates for fungal recognition receptors. Much sequence homology is present between these C-type lectins and mammalian Dectin-1, Dectin-2, Mincle, and DC-SIGN, and some syntenic relationships have been identified. Although phylogenetic studies indicate that Group V C-type lectins

are absent in fish, it is still possible that zebrafish possess receptors that recognize pathogenic fungi.

In the future, it would be ideal to fully characterize the ligand binding properties of zSCLRA, zKCLectin, and zDC-SIGN-like F through the use of FRR-Fc fusion proteins, expression in HEK293T cells, or morpholino technology. If none of techniques identify a zebrafish receptor for fungi, it may be of use to revisit the Group II C-type lectin database, identify other potential receptors, and to clone and express these receptors in the same manner as for the current three candidate receptors. Alternatively, one could inject zebrafish with beta-glucan extracts and monitor the immune response or use morpholinos to knock down intermediates in the signal transduction pathway induced by Dectin-1.

Determining whether or not zebrafish possess fungal recognition receptors is important in enhancing our understanding of how the innate immune system has evolved in lower and higher vertebrates. Understanding such dynamics is a key step in determining how fungi universally infect all organisms and may provide additional insights in several areas of research, such as developing anti-fungal therapies for commercially valuable fish. We hope that this work opens doors for candidiasis research and helps uncover the fundamental mechanisms of vertebrate immunity to fungi.

WORKS CITED


1. http://www.ncbi.nlm.nih.gov/gene?term=chicken
2. http://www.ncbi.nlm.nih.gov/gene?term=playtpus
3. 2002. Phylogenetic Inference. In *Encyclopedia of Evolution*. M. Pagel, ed. Oxford University Press, New York, N.Y.
4. Adachi, Y., T. Ishii, Y. Ikeda, A. Hoshino, H. Tamura, J. Aketagawa, S. Tanaka, and N. Ohno. 2004. Characterization of beta-glucan recognition site on C-type lectin, dectin 1. *Infect Immun* 72:4159-4171.
5. Ariizumi, K., G. L. Shen, S. Shikano, R. Ritter, 3rd, P. Zukas, D. Edelbaum, A. Morita, and A. Takashima. 2000. Cloning of a second dendritic cell-associated C-type lectin (dectin-2) and its alternatively spliced isoforms. *J Biol Chem* 275:11957-11963.
6. Bigas, A., A. Robert-Moreno, and L. Espinosa. 2010. The Notch pathway in the developing hematopoietic system. *Int J Dev Biol* 54:1175-1188.
7. Bourque, G., and G. Tesler. 2008. Computational tools for the analysis of rearrangements in mammalian genomes. *Methods Mol Biol* 452:431-455.
8. Brown, G. D., D. W. Denning, and S. M. Levitz. 2012. Tackling human fungal infections. *Science* 336:647.
9. Brown, J., C. A. O'Callaghan, A. S. Marshall, R. J. Gilbert, C. Siebold, S. Gordon, G. D. Brown, and E. Y. Jones. 2007. Structure of the fungal beta-glucan-binding immune receptor dectin-1: implications for function. *Protein Sci* 16:1042-1052.
10. Catchen, J. M., J. S. Conery, and J. H. Postlethwait. 2008. Inferring ancestral gene order. *Methods Mol Biol* 452:365-383.
11. Catchen, J. M., J. S. Conery, and J. H. Postlethwait. 2009. Automated identification of conserved synteny after whole-genome duplication. *Genome Res* 19:1497-1505.
12. De Lucca, A. J. 2007. Harmful fungi in both agriculture and medicine. *Rev Iberoam Micol* 24:3-13.
13. Fisher, M. C., D.A. Henk, C.J. Briggs, J.S. Brownstein, L.C. Madoff, S.L. McCraw, and S.J. Gurr. 2012. Emerging fungal threats to animal, plant and ecosystem health. *Nature* 484:186-194.
14. Fuller, G. L., J. A. Williams, M. G. Tomlinson, J. A. Eble, S. L. Hanna, S. Pohlmann, K. Suzuki-Inoue, Y. Ozaki, S. P. Watson, and A. C. Pearce. 2007. The C-type lectin receptors CLEC-2 and Dectin-1, but not DC-SIGN, signal via a novel YXXL-dependent signaling cascade. *J Biol Chem* 282:12397-12409.
15. Gardiner, K. 2004. Synteny. In *Dictionary of Bioinformatics and Computational Biology*. J. M. Hancock, and M.J. Zvelebil, ed. John Wiley & Sons, Hoboken, N.J.
16. Graham, L. M., and G. D. Brown. 2009. The Dectin-2 family of C-type lectins in immunity and homeostasis. *Cytokine* 48:148-155.
17. Heringa, J. 2004. Phylogenetic Tree (Phylogenetic Reconstruction, Phylogeny, Phylogeny Reconstruction). In *Dictionary of Bioinformatics and Computational Biology*. J. M. Hancock, and M.J. Zvelebil, ed. John Wiley & Sons, Hoboken, N.J.

18. Herrero, A. B., P. Magnelli, M. K. Mansour, S. M. Levitz, H. Bussey, and C. Abeijon. 2004. KRE5 gene null mutant strains of Candida albicans are avirulent and have altered cell wall composition and hypha formation properties. *Eukaryot Cell* 3:1423-1432.
19. Hollmig, S. T., K. Ariizumi, and P. D. Cruz, Jr. 2009. Recognition of non-self-polysaccharides by C-type lectin receptors dectin-1 and dectin-2. *Glycobiology* 19:568-575.
20. Hu, Y. L., X. M. Pan, L. X. Xiang, and J. Z. Shao. 2010. Characterization of C1q in teleosts: insight into the molecular and functional evolution of C1q family and classical pathway. *J Biol Chem* 285:28777-28786.
21. Jackson, A. N., C. A. McLure, R. L. Dawkins, and P. J. Keating. 2007. Mannose binding lectin (MBL) copy number polymorphism in Zebrafish (D. rerio) and identification of haplotypes resistant to L. anguillarum. *Immunogenetics* 59:861-872.
22. Julenius, K., and A. G. Pedersen. 2006. Protein evolution is faster outside the cell. *Mol Biol Evol* 23:2039-2048.
23. Kaneda, M., T. Odaka, H. Suetake, D. Tahara, and T. Miyadai. 2012. Teleost IL-6 promotes antibody production through STAT3 signaling via IL-6R and gp130. *Dev Comp Immunol* 38:224-231.
24. Kindt, T. J., R. A. Goldsby, and B. A. Osborne. 2007. *Kuby Immunology*. W. H. Freeman and Company, New York.
25. Kumar, S., and A. Filipski. 2004. Maximum-Likelihood Phylogeny Reconstruction. In *Dictionary of Bioinformatics and Computational Biology*. J. M. Hancock, and M.J. Zvelebil, ed. John Wiley & Sons, Hoboken, N.J.
26. Kumar, S., and A. Filipski. 2004. Neighbor-Joining Method. In *Dictionary of Bioinformatics and Computational Biology*. J. M. Hancock, and M.J. Zvelebil, ed. John Wiley & Sons, Hoboken, N.J.
27. Liao, B. Y., M. P. Weng, and J. Zhang. 2010. Contrasting genetic paths to morphological and physiological evolution. *Proc Natl Acad Sci U S A* 107:7353-7358.
28. Lin, A. F., L. X. Xiang, Q. L. Wang, W. R. Dong, Y. F. Gong, and J. Z. Shao. 2009. The DC-SIGN of zebrafish: insights into the existence of a CD209 homologue in a lower vertebrate and its involvement in adaptive immunity. *J Immunol* 183:7398-7410.
29. Luz, H., and M. Vingron. 2006. Family specific rates of protein evolution. *Bioinformatics* 22:1166-1171.
30. McGreal, E. P., M. Rosas, G.D. Brown, S. Zamze, S.Y.C. Wong, S. Gordon, L. Martinez-Pomares, and P.R. Taylor. 2005. The carbohydrate-recognition domain of Dectin-2 is a C-type lectin with specificity for high mannose. *Glycobiology* 16:422-430.
31. Nobrega, M. A., and L.A. Pennacchio. 2003. Comparative genomic analysis as a tool for biological discovery. *J Physiol* 554:31-39.
32. Osorio, F., and C. Reis e Sousa. 2011. Myeloid C-type lectin receptors in pathogen recognition and host defense. *Immunity* 34:651-664.
33. Panagos, P. G., K. P. Dobrinski, X. Chen, A. W. Grant, D. Traver, J. Y. Djeu, S. Wei, and J. A. Yoder. 2006. Immune-related, lectin-like receptors are

differentially expressed in the myeloid and lymphoid lineages of zebrafish. *Immunogenetics* 58:31-40.
34. Pfaller, M. A., and D. J. Diekema. 2007. Epidemiology of invasive candidiasis: a persistent public health problem. *Clin Microbiol Rev* 20:133-163.
35. Postlethwait, J. H., Y. L. Yan, M. A. Gates, S. Horne, A. Amores, A. Brownlie, A. Donovan, E. S. Egan, A. Force, Z. Gong, C. Goutel, A. Fritz, R. Kelsh, E. Knapik, E. Liao, B. Paw, D. Ransom, A. Singer, M. Thomson, T. S. Abduljabbar, P. Yelick, D. Beier, J. S. Joly, D. Larhammar, F. Rosa, M. Westerfield, L. I. Zon, S. L. Johnson, and W. S. Talbot. 1998. Vertebrate genome evolution and the zebrafish gene map. *Nat Genet* 18:345-349.
36. Pyz, E., A. S. Marshall, S. Gordon, and G. D. Brown. 2006. C-type lectin-like receptors on myeloid cells. *Ann Med* 38:242-251.
37. Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425.
38. Sato, K., X. L. Yang, T. Yudate, J. S. Chung, J. Wu, K. Luby-Phelps, R. P. Kimberly, D. Underhill, P. D. Cruz, Jr., and K. Ariizumi. 2006. Dectin-2 is a pattern recognition receptor for fungi that couples with the Fc receptor gamma chain to induce innate immune responses. *J Biol Chem* 281:38854-38866.
39. Sisson, T. H., and C. W. Castor. 1990. An improved method for immobilizing IgG antibodies on protein A-agarose. *J Immunol Methods* 127:215-220.
40. Soanes, K. H., K. Figuereido, R. C. Richards, N. R. Mattatall, and K. V. Ewart. 2004. Sequence and expression of C-type lectin receptors in Atlantic salmon (Salmo salar). *Immunogenetics* 56:572-584.
41. Spence, R., G. Gerlach, C. Lawrence, and C. Smith. 2008. The behaviour and ecology of the zebrafish, Danio rerio. *Biol Rev Camb Philos Soc* 83:13-34.
42. Steele, C., R. R. Rapaka, A. Metz, S. M. Pop, D. L. Williams, S. Gordon, J. K. Kolls, and G. D. Brown. 2005. The beta-glucan receptor dectin-1 recognizes specific morphologies of Aspergillus fumigatus. *PLoS Pathog* 1:e42.
43. Sunyer, J. O. 2013. Fishing for mammalian paradigms in the teleost immune system. *Nat Immunol* 14:320-326.
44. Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731-2739.
45. Tesler, G. 2002. GRIMM: genome rearrangements web server. *Bioinformatics* 18:492-493.
46. Wang, Z., S. Zhang, and G. Wang. 2008. Response of complement expression to challenge with lipopolysaccharide in embryos/larvae of zebrafish Danio rerio: acquisition of immunocompetent complement. *Fish Shellfish Immunol* 25:264-270.
47. Wheeler, R. T., D. Kombe, S. D. Agarwala, and G. R. Fink. 2008. Dynamic, morphotype-specific Candida albicans beta-glucan exposure during infection and drug treatment. *PLoS Pathog* 4:e1000227.
48. White, T. C., S. Holleman, F. Dy, L. F. Mirels, and D. A. Stevens. 2002. Resistance mechanisms in clinical isolates of Candida albicans. *Antimicrob Agents Chemother* 46:1704-1713.

49. Willment, J. A., S. Gordon, and G. D. Brown. 2001. Characterization of the human beta -glucan receptor and its alternatively spliced isoforms. *J Biol Chem* 276:43818-43823.
50. Wilson, C. W., and P. T. Chuang. 2010. Mechanism and evolution of cytosolic Hedgehog signal transduction. *Development* 137:2079-2094.
51. Woods, I. G., P. D. Kelly, F. Chu, P. Ngo-Hazelett, Y. L. Yan, H. Huang, J. H. Postlethwait, and W. S. Talbot. 2000. A comparative map of the zebrafish genome. *Genome Res* 10:1903-1914.
52. Yang, J. 2009. Immune-related Lectin-Like Receptors in Zebrafish Innate Immunity. North Carolina State University, Raleigh, N.C. 138.
53. Zelensky, A. N., and J. E. Gready. 2004. C-type lectin-like domains in Fugu rubripes. *BMC Genomics* 5:51.
54. Zelensky, A. N., and J. E. Gready. 2005. The C-type lectin-like domain superfamily. *FEBS J* 272:6179-6217.
55. Zimmerman, A. M., F. M. Moustafa, K. E. Romanowski, and L. A. Steiner. 2011. Zebrafish immunoglobulin IgD: unusual exon usage and quantitative expression profiles with IgM and IgZ/T heavy chain isotypes. *Mol Immunol* 48:2220-2223.
56. Zimmerman, A. M., G. Yeo, K. Howe, B. J. Maddox, and L. A. Steiner. 2008. Immunoglobulin light chain (IgL) genes in zebrafish: Genomic configurations and inversional rearrangements between (V(L)-J(L)-C(L)) gene clusters. *Dev Comp Immunol* 32:421-434.

TABLES

**Table 1. GenBank reference numbers of alignment and phylogenetic sequences**

| Lectin Name | Organism | Ref. Number or Paper |
|---|---|---|
| ASGPR2 | *D. rerio* | XP_001339645.1 |
| CD209 | *D. rerio* | ADB55613.1 |
| DC-SIGN-like A | *D. rerio* | XP_002661262.1 |
| DC-SIGN-like B | *D. rerio* | NP_001107113.1 |
| DC-SIGN-like C | *D. rerio* | XP_002663334.2 |
| DC-SIGN-like D | *D. rerio* | XP_003199386.1 |
| DC-SIGN-like E | *D. rerio* | XP_003199401.1 |
| DC-SIGN-like F | *D. rerio* | NP_001170922.1 |
| DC-SIGN-like G | *D. rerio* | XP_003197805.1 |
| DC-SIGN-like H | *D. rerio* | XP_001919882.2 |
| DC-SIGN-like I | *D. rerio* | XP_003199385.1 |
| Illr1 | *D. rerio* | NP_001035139.1 |
| Illr2 | *D. rerio* | NP_001121843.1 |
| Illr3a | *D. rerio* | NP_001035128.1 |
| Illr3b | *D. rerio* | NP_001038207.1 |
| Illr4 | *D. rerio* | NP_001035129.1 |
| KCLectin | *D. rerio* | XP_001335195.1 |
| KCLR | *D. rerio* | XP_003197803.1 |
| Mincle | *D. rerio* | XP_002660626.2 |
| SCLRA | *D. rerio* | NP_001038504.2 |
| Unknown CTLD B | *D. rerio* | XP_003198053.1 |
| Unknown CTLD C | *D. rerio* | XP_003199381.1 |
| Unknown CTLD I | *D. rerio* | XP_001919882.2 |
| CD69 | *Gallus gallus* | NP_001074183.1 |
| CD72 | *Gallus gallus* | NP_990383.1 |
| KCLectinG | *Gallus gallus* | XP_003643020.1 |
| NKR | *Gallus gallus* | NP_001038147.1 |
| Unknown1 | *Gallus gallus* | NP_990760.1 |
| ASGPR1 | *H. sapiens* | NP_001662.1 |
| ASGPR2 | *H. sapiens* | NP_001172.1 |
| CD209 | *H. sapiens* | NP_001138371.1 |
| CD69 | *H. sapiens* | NP_001772.1 |
| CD72 | *H. sapiens* | NP_001773.1 |
| CD94 | *H. sapiens* | NP_001107868.1 |
| CLAX | *H. sapiens* | NP_037401.1 |
| DCIR | *H. sapiens* | NP_057268.1 |
| DC-SIGNR | *H. sapiens* | AAI10615.1 |
| Dectin-1 | *H. sapiens* | NP_922938.1 |

**Table 1. (continued)**

| Lectin Name | Organism | Ref. Number or Paper |
|---|---|---|
| Dectin-2 | *H. sapiens* | NP_001007034.1 |
| HML2 | *H. sapiens* | NP_878910.1 |
| KCLR | *H. sapiens* | NP_775806.2 |
| MCL | *H. sapiens* | NP_525126.2 |
| MDL1 | *H. sapiens* | NP_037384.1 |
| Mincle | *H. sapiens* | NP_055173.1 |
| NKG2A | *H. sapiens* | NP_002250.1 |
| NKG2D | *H. sapiens* | NP_031386.2 |
| ORL1 | *H. sapiens* | NP_002534.1 |
| ASGPR1 | *M. musculus* | NP_033844. |
| ASGPR2 | *M. musculus* | NP_031519.1 |
| CD209a | *M. musculus* | NP_573501.1 |
| CD69 | *M. musculus* | NP_001028294.1 |
| CD72 | *M. musculus* | NP_001103790.1 |
| CD94 | *M. musculus* | NP_034784.1 |
| DCIR | *M. musculus* | NP_001163803.1 |
| Dectin-1 | *M. musculus* | NP_064392.2 |
| Dectin-2 | *M. musculus* | NP_001177249.1 |
| MCL | *M. musculus* | NP_001156633.1 |
| MDL | *M. musculus* | NP_001033693.1 |
| Mincle | *M. musculus* | NP_064332.1 |
| NKG2A | *M. musculus* | NP_001129540.1 |
| NKG2D | *M. musculus* | NP_149069.1 |
| OCIL | *M. musculus* | NP_444339.1 |
| OLR1 | *M. musculus* | NP_619589.2 |
| SCLRA | *S. salar* | AAT77220.1 |
| SCLRB | *S. salar* | AAT77221.1 |
| SCLRC | *S. salar* | AAT77222.1 |
| FDC-SIGN-F1 | *T. rubripes* | Zelensky 2004 |
| FDC-SIGN-F2 | *T. rubripes* | Zelensky 2004 |
| FDC-SIGN-F3 | *T. rubripes* | Zelensky 2004 |
| FDC-SIGN-F4 | *T. rubripes* | Zelensky 2004 |
| FDC-SIGN-F5 | *T. rubripes* | Zelensky 2004 |
| FDC-SIGN-F6 | *T. rubripes* | Zelensky 2004 |
| FDC-SIGN-F7 | *T. rubripes* | Zelensky 2004 |
| FDC-SIGN-F8 | *T. rubripes* | Zelensky 2004 |
| FDC-SIGNR | *T. rubripes* | Zelensky 2004 |

**Table 2. List of CRD Primers**

| Primer Name | Sequence | Length (bp) |
|---|---|---|
| zSCLRA Fwd | AGGTGGTACCGATGTCCAGAAGATTGGATGC | 31 |
| zSCLRA Rev | AGGGAATTCGACACTAAAGGCGATAGGGTC | 30 |
| zKC-Lectin Fwd | AGGTGGTACCGATGTGGTAGTGGATGGCTGAG | 32 |
| zKC-Lectin Rev | AGGGAATTCCATCTTCTGACAAATGTACTG | 30 |
| zDC-SIGN-like F Fwd | AGGTGGTACCGATGTCCTCAAAACTGGAAGCAC | 33 |
| zDC-SIGN-like F Rev | AGGGAATTCAGAGGATGCCGGTTTCTCAC | 29 |

**Table 3. Conserved amino acid motifs in vertebrate C-type lectin fungal recognition receptors**

| Amino Acid Sequence | Function | Sequences Conserved In |
|---|---|---|
| YXXL | Immunoreceptor tyrosine activation-like motif. Initiates signal transduction. | Dectin-1, zKCLectin, zSCLRA, sSCLRA |
| R | Transmembrane or cytosolic arginine. Couples to FcRγ to initiate signal transduction. | Dectin-2, Mincle |
| LL | Double leucine repeat | CD209, DC-SIGNR, zCD209 |
| EE (E/D) | Tri-acid cluster | CD209, DC-SIGNR, zCD209, zDC-SIGN-like F |
| EPN | Mannose-recognition sequence, calcium binding site | Dectin-2, Mincle, DC-SIGNR, CD209, zCD209, zDC-SIGN-like F |
| WIGL | Carbohydrate recognition site | zCD209, zKCLectin, zSCLRA, Dectin-1, sSCLRA |
| WND | Calcium binding site | Dectin-2, Mincle, CD209, DC-SIGNR |
| WIH | Beta-glucan reocngition sequence | Dectin-1, sSCLRA (maybe) |

**Table 4a. Percent homology between zebrafish and human syntenic genes**

| Human Protein | Zebrafish Protein | % Homology |
|---|---|---|
| Dectin-1 | zDC-SIGN-like F | 32 |
| Dectin-1 | zSCLRA | 60 |
| Dectin-1 | zKCLectin | 23 |
| Dectin-2 | zDC-SIGN-like F | 34 |
| Dectin-2 | zSCLRA | 29 |
| Dectin-2 | zKCLectin | 27 |
| Mincle | zDC-SIGN-like F | 41 |
| Mincle | zSCLRA | 39 |
| Mincle | zKCLectin | 28 |
| CD209 | zDC-SIGN-like F | 50 |
| CD209 | zSCLRA | 32 |
| CD209 | zKCLectin | 63 |
| GABARAPL1 | GABARAP | 59 |
| COPS7A | COPS7A | 57 |
| ZNF384 | ZNF384 | 41 |
| PEX5 | PEX5 | 60 |
| CLSTN3 | CLSTN3 | 60 |
| LPCAT3 | LPCAT3 | 64 |
| TPI1 | TPI1 | 84 |
| GNB3 | GNB3 | 80 |

**Table 4b. Percent homology between mouse and zebrafish syntenic genes**

| Mouse Protein | Zebrafish Protein | % Homology |
|---|---|---|
| Dectin-1 | zDC-SIGN-like F | 29 |
| Dectin-1 | zSCLRA | 27 |
| Dectin-1 | zKCLectin | 38 |
| Dectin-2 | zDC-SIGN-like F | 43 |
| Dectin-2 | zSCLRA | 28 |
| Dectin-2 | zKCLectin | 27 |
| Mincle | zDC-SIGN-like F | 44 |
| Mincle | zSCLRA | 36 |
| Mincle | zKCLectin | 28 |
| CD209 | zDC-SIGN-like F | 36 |
| CD209 | zSCLRA | 29 |
| CD209 | zKCLectin | 31 |
| GABARAPL1 | GABARAP | 59 |
| COPS7A | COPS7A | 58 |
| ZNF384 | ZNF384 | 56 |
| PEX5 | PEX5 | 58 |
| CLSTN3 | CLSTN3 | 61 |
| LPCAT3 | LPCAT3 | 64 |
| TPI1 | TPI1 | 81 |
| GNB3 | GNB3 | 79 |

**Table 4c. Percent homology between human and mouse syntenic genes**

| Human Protein | Mouse Protein | % Homology |
|---|---|---|
| Dectin-1 | Dectin-1 | 61 |
| Dectin-2 | Dectin-2 | 68 |
| Mincle | Mincle | 67 |
| CD209 | CD209 | 56 |
| GABARAPL1 | GABARAPL1 | 100 |
| COPS7A | COPS7A | 99 |
| ZNF384 | ZNF384 | 85 |
| PEX5 | PEX5 | 91 |
| CLSTN3 | CLSTN3 | 96 |
| LPCAT3 | LPCAT3 | 87 |
| TPI1 | TPI1 | 96 |
| GNB3 | GNB3 | 97 |

FIGURES

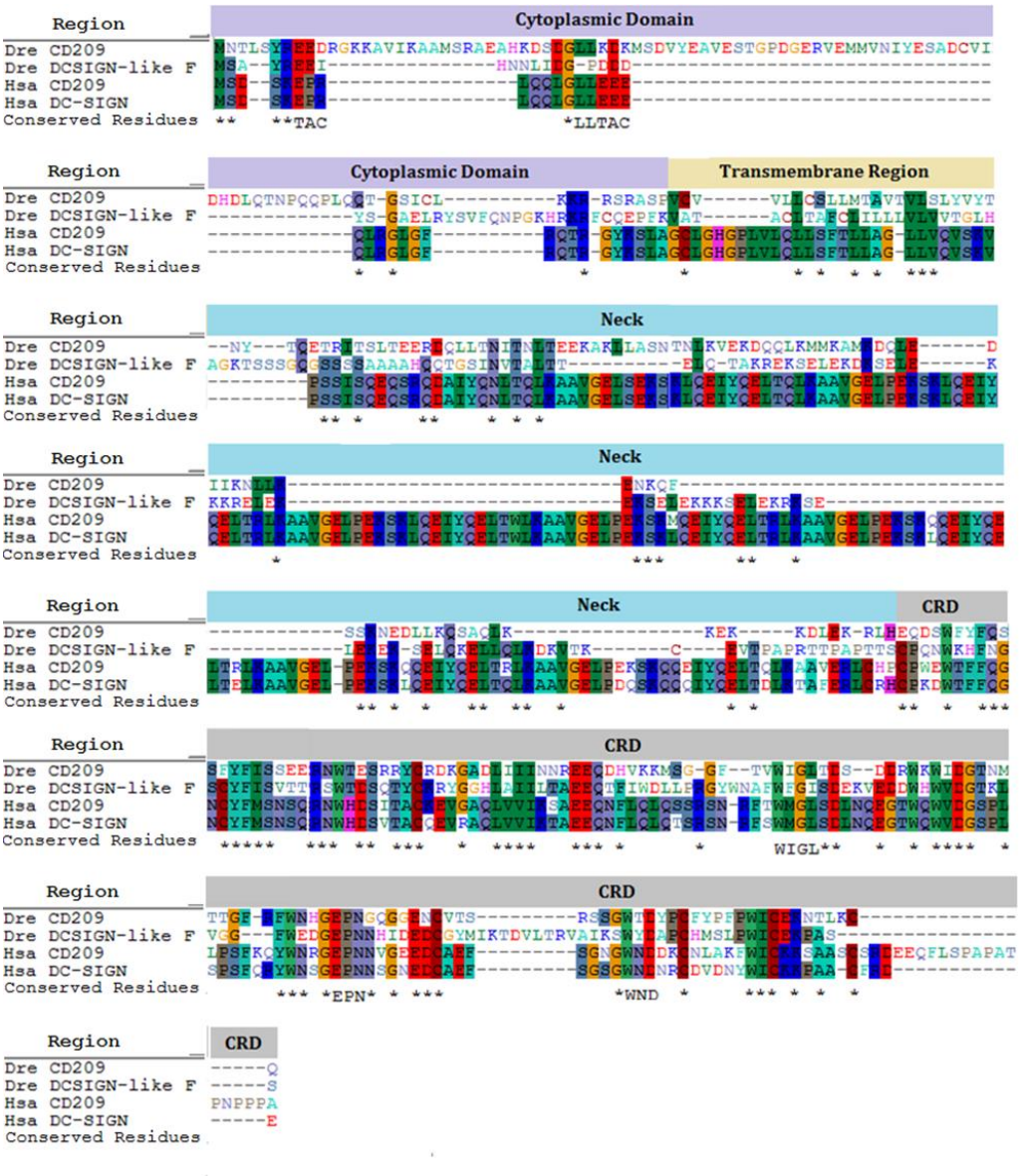


**Figure 1**. **The carbohydrate recognition domain (CRD) of zebrafish DC-SIGN-like F shares homology with zebrafish CD209, human CD209 and human DC-SIGN (CLEC4M).** Full amino acid sequences of zCD209, zDC-SIGN-like F, human CD209, and human DC-SIGN (GenBank access numbers ADB55613.1, NP_001170922.1, NP_001138371.1, AAI10615.1) were aligned with ClustalW and manually edited in BioEdit. The cytoplasmic, transmembrane, central stalk, and carbohydrate recognition domain (CRD) regions are outlined above the sequences with colored boxes. Amino acid letters are colored based on their chemical properties. Highlighted amino acids are colored similarly. Conserved residues are indicated under the alignment with asterisk. Important structural sequences, including specific carbohydrate and calcium recognition sequences (WIGL, EPN, WND) are indicated within the sequences and under conserved residues. The cytoplasmic domain and neck regions of zCD209 differ from those of the other proteins. The neck region of zDC-SIGN-like F contains multiple repeats, but to a lesser extent than human CD209 and human DC-SIGN.

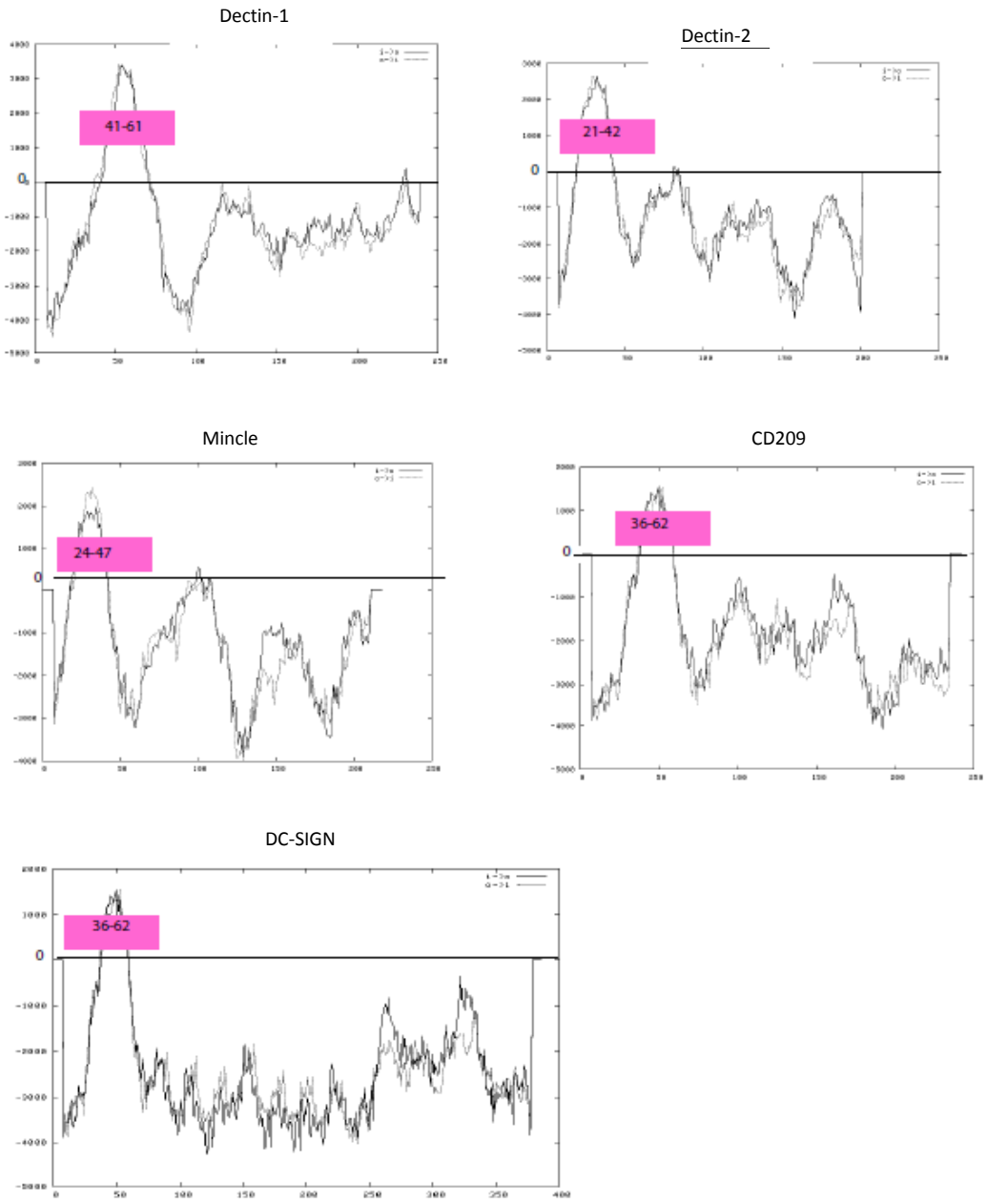


**Figure 2. The cytoplasmic and extracellular domains of human and potential zebrafish fungal recognition receptors vary in size.** A.) Membrane topologies for mammalian amino acid sequences of interest were predicted using TMPRED (http://www.ch.embnet.org/software/TMPRED_form.html). B.) Membrane topologies for fish amino acid sequences of interest were predicted using TMPRED. For both sets, X-axis values indicate amino acid number. Y-axis values greater than zero over a range of 17 to 33 amino acids indicate hydrophobic transmembrane regions. The data from TMPRED as well as TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) was used to construct Figures 1 and 4.

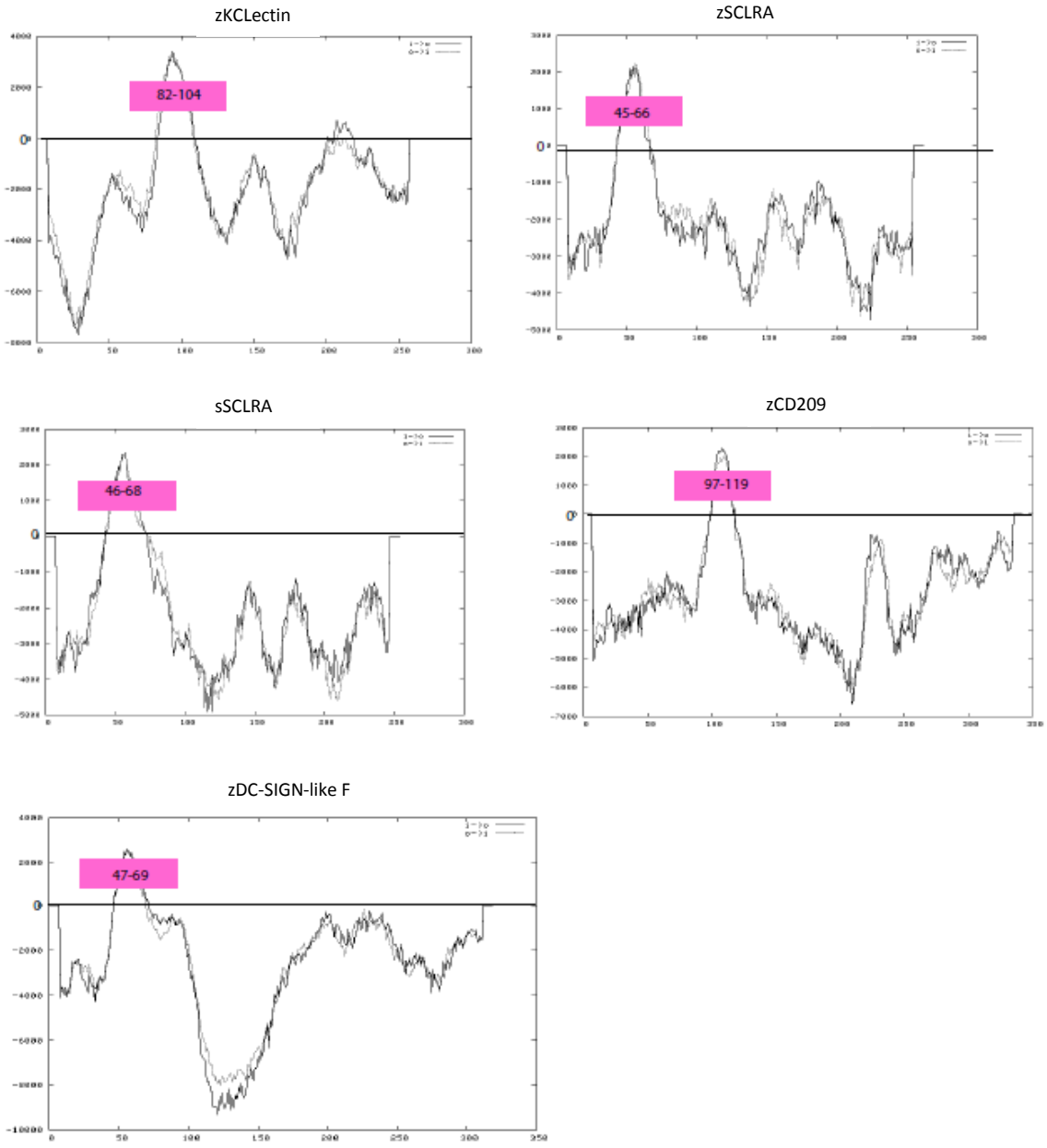


**Figure 3. The cytoplasmic and extracellular domains of human and potential zebrafish fungal recognition receptors vary in size.** A.) Membrane topologies for mammalian amino acid sequences of interest were predicted using TMPRED (http://www.ch.embnet.org/software/TMPRED_form.html). B.) Membrane topologies for fish amino acid sequences of interest were predicted using TMPRED. For both sets, X-axis values indicate amino acid number. Y-axis values greater than zero over a range of 17 to 33 amino acids indicate hydrophobic transmembrane regions. The data from TMPRED as well as TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) was used to construct Figures 1 and 4.

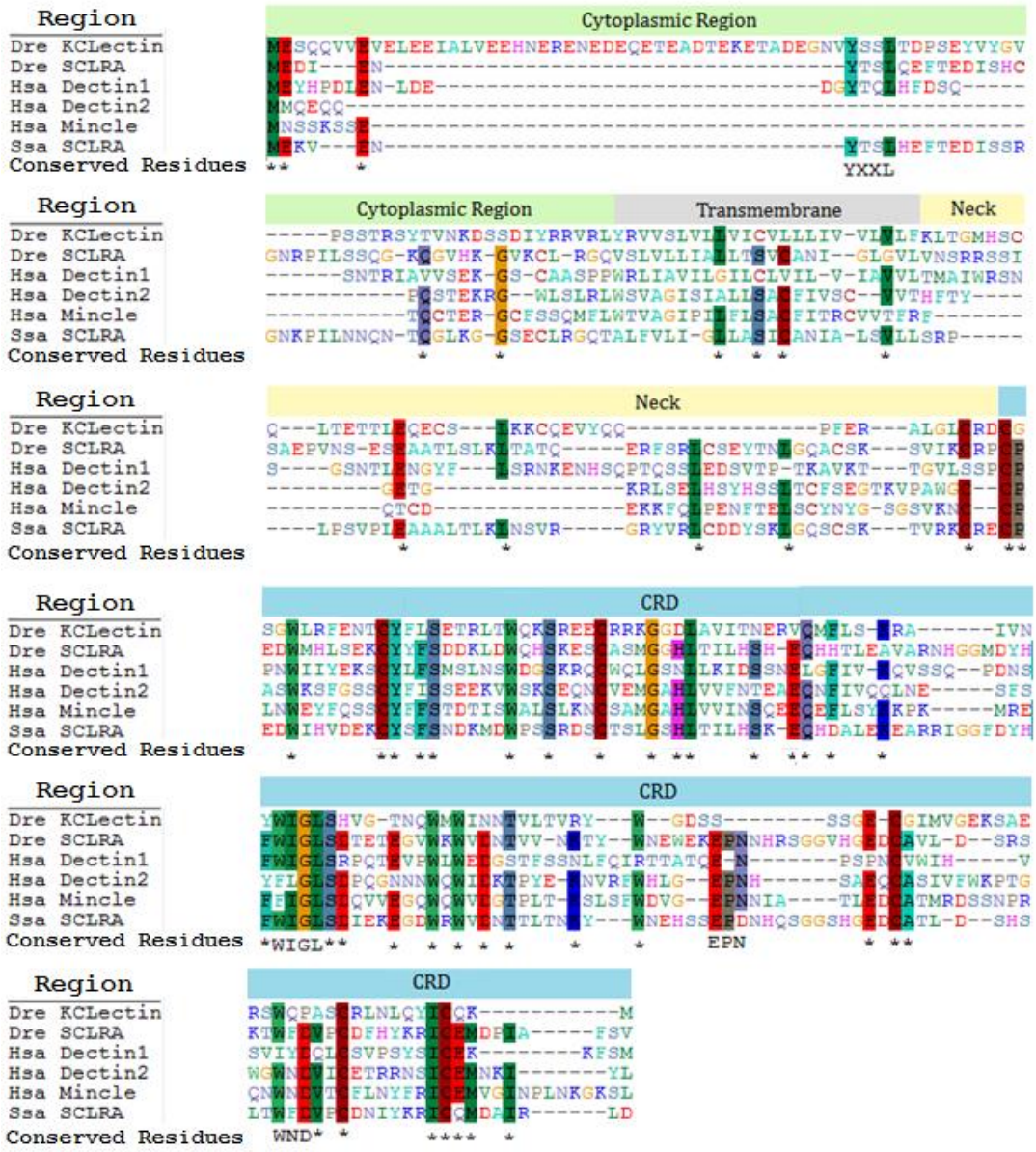


**Figure 4. The mammalian CRD region and immunoreceptor tyrosine activation motifs (ITAMs) are conserved in zKCLectin and zSCLRA.** Full amino acid sequences of zebrafish SCLRA, KCLectin, and salmon SCLRA (GenBank access numbers NP_001038504.2, XP_001335195.1, AAT77220.1) were aligned with human Dectin-1, Dectin-2, and Mincle (GenBank access numbers NP_922938.1, NP_001007034.1, NP_055173.1) using ClustalW. Alignments were manually edited in BioEdit. The cytoplasmic, transmembrane, central stalk, and carbohydrate recognition domain (CRD) regions are outlined above the sequences with colored boxes. Conserved residues are indicated under Conserved Residues. Amino acid letters are colored based on their chemical properties. Highlighted amino acids are colored similarly. Conserved residues are indicated under the alignment with asterisk. Important structural sequences, including specific carbohydrate and calcium recognition sequences (WIGL, EPN, WND) and potential hemi-ITAMs are indicated within the sequences and under conserved residues. The cytoplasmic region of zKCLectin diverges from other cytoplasmic regions. Variation also exists within the neck region.

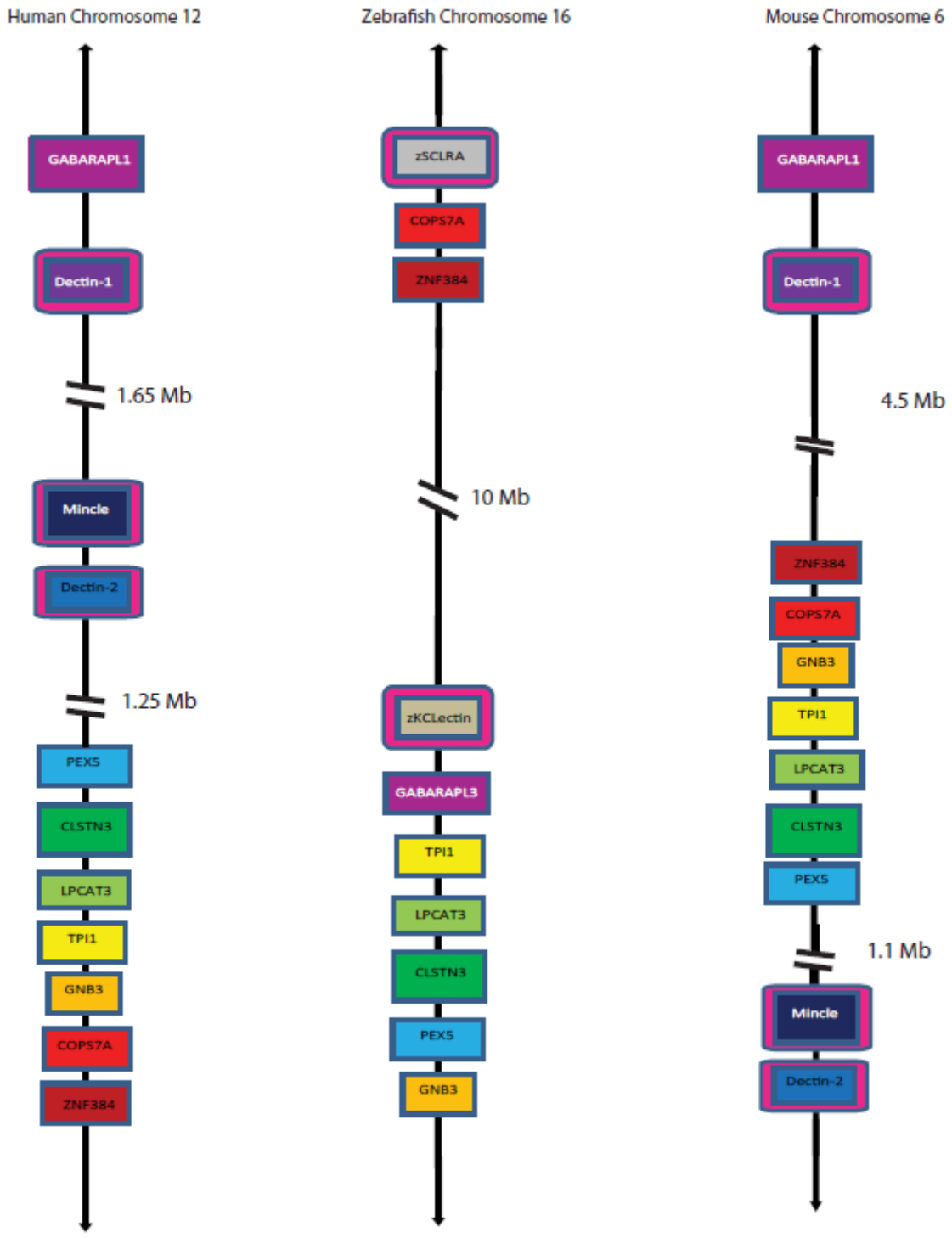


**Figure 5. Synteny is present between potential zebrafish fungal recognition receptors (FRRs) and the mammalian NK cluster.** Relationships between zebrafish KCLectin and SCLRA and mammalian FRRs. Potential homology of zebrafish KCLectin and SCLRA with Dectin-1, Dectin-2, and Mincle is highlighted by a pink box. Significant gaps in synteny between species are indicated and the chromosomal representations are schematics and not drawn to scale. CTLDs are highlighted in pink boxes. Genomic locations for human chromosome 12, zebrafish chromosome 16, and mouse chromosome 6 were obtained from the National Center of Biotechnology Information and Ensembl databases.

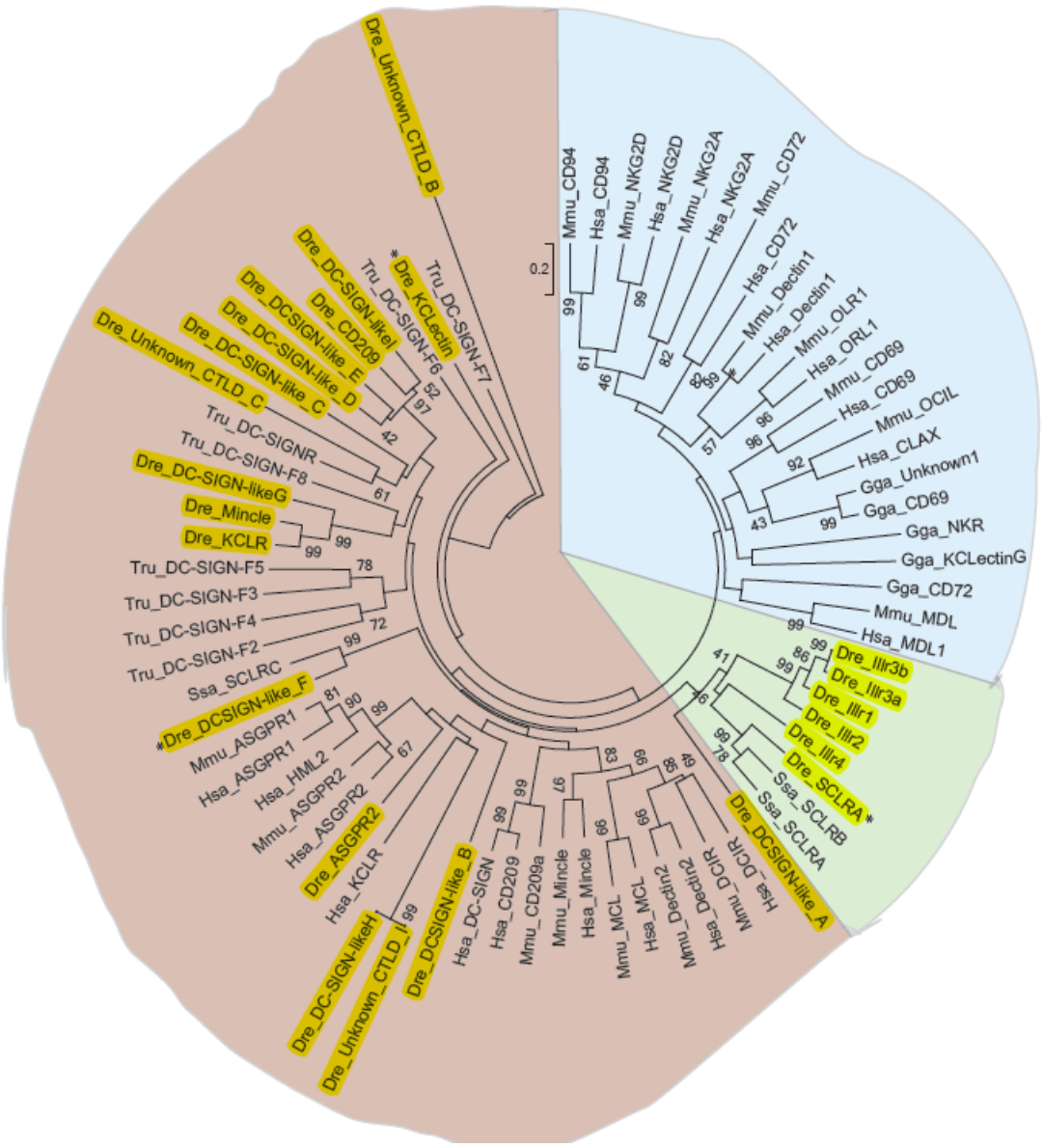


**Figure 6. Mammalian Group V C-type lectin receptors and teleost NK-like receptors evolved independently from Group II C-type lectin receptors.** Non-redundant CTLD sequences from Group II (red), Group V (blue), and NK-like (green) C-type lectins in *D. rerio*, *T. rubripes*, *S. salar*, *G. gallus*, *M. musculus*, and *H. sapiens* were aligned by ClustalW. A radial neighbor-joining tree was generated by MEGA5 using the Poisson distribution and 2000 bootstrap replications. Branch lengths are measured in terms of amino acid substitutions, with the scale indicated next to *M. musculus* CD94. Protein names are preceded by species symbols (Dre *Danio rerio*, Tru *Takifugu rubripes*, Ssa *Salmo salar*, Gga *Galllus gallus*, Mmu *Mus musculus*, Hsa *Homo sapiens). D. rerio* sequences are highlighted in yellow. Lectins that were later cloned are indicated by an asterisk. Bootstrap values exceeding 40% are indicated.

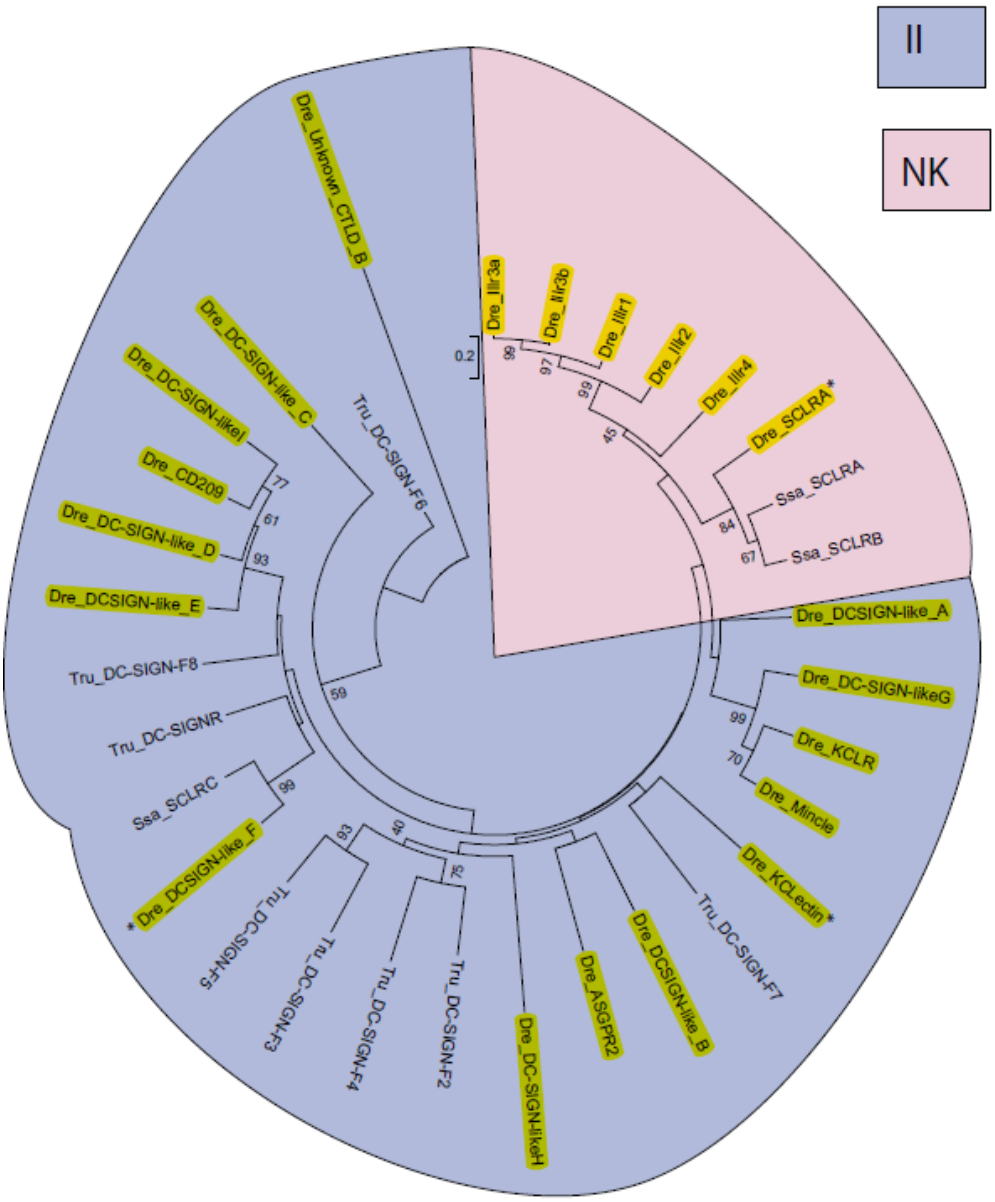


**Figure 7. NK-like receptors evolved independently in teleost fish.** Non-redundant CTLD sequences from *D. rerio*, *T. rubripes*, and *S. salar* Group II (purple) and NK-like (pink) C-type lectins were aligned by ClustalW. A radial neighbor-joining tree was generated by MEGA5 using the Poisson distribution and 2000 bootstrap replications. Branch lengths are measured in terms of amino acid substitutions, with the scale indicated next to *D. rerio* Illr3a. Protein names are preceded by species symbols (Dre *Danio rerio*, Tru *Takifugu rubripes*, Ssa *Salmo salar*). *D. rerio* sequences are highlighted in yellow. Lectins that were later cloned are indicated by an asterisk. Bootstrap values exceeding 40% are indicated.

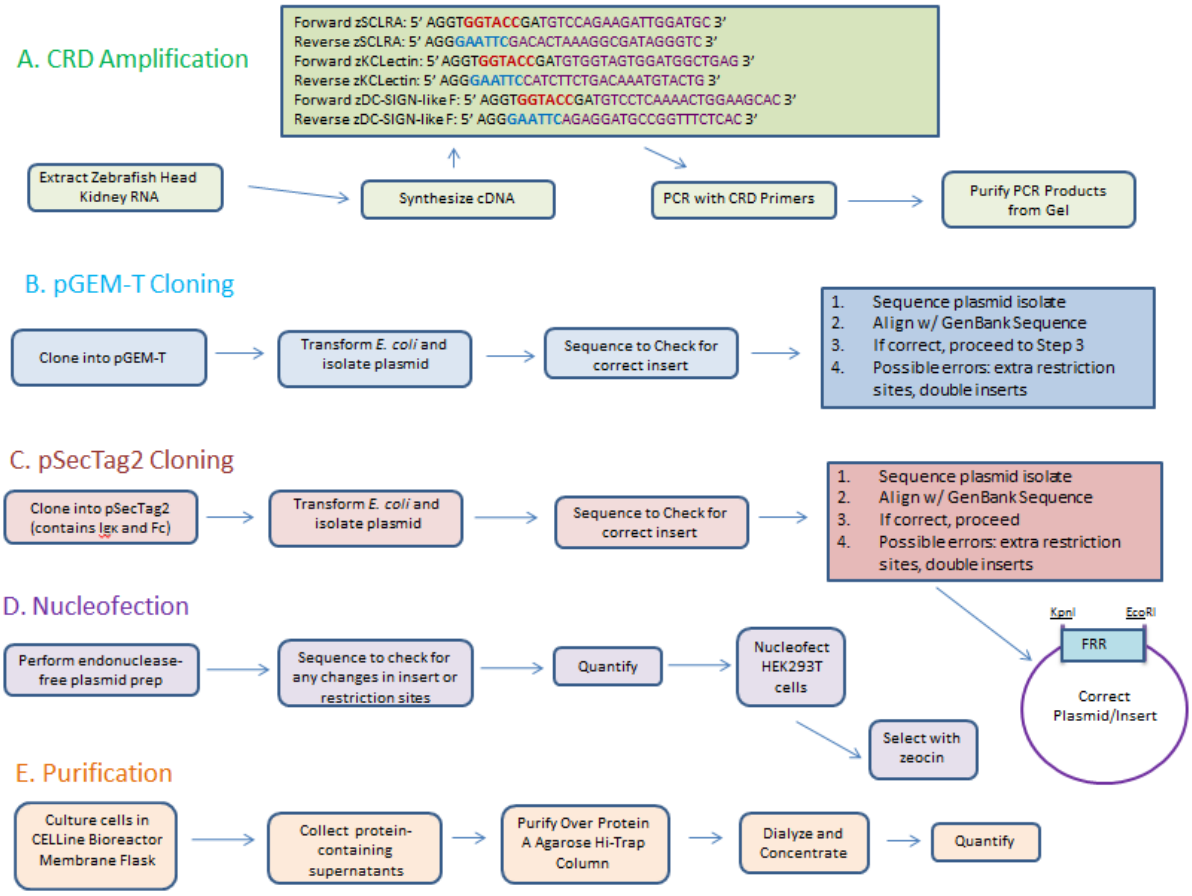


**Figure 8. Expression and Purification of Fungal Recognition Receptor CRDs.**
A.) CRDs for zSCLRA, zKCLectin, and zDC-SIGN-like F were amplified from zebrafish head kidney cDNA. For the PCR primers, the KpnI restriction enzyme site is outlined in red, and the EcoRI restriction enzyme site is outlined in blue. B.) Amplicons were cloned into pGEM-T and sequenced to check for the correct insert. C.) CRD amplicons in pGEM-T were subcloned into pSecTag2 and sequenced. D.) FRR-pSecTag2 vectors were nucleofected into HEK293T cells after an endonuclease-free plasmid preparation and selected with 100 µg/ml Zeocin to establish stable lines. E.) Supernatants containing fusion protein were harvested from a CELLline Bioreactor Membrane Flask. The protein was purified over protein A agarose, and quantified with a Bradford Assay.

ELECTRONIC APPENDIX

| Name | File Type | Section, Page | File Path |
|---|---|---|---|
| All Zebrafish CTLD Genes | xls | Materials and Methods, 14 | file:///\\TS-QVHLC90\share\Erin C\Thesis Electronic Appendix\All Zebrafish CTLD Genes.xls |
| Group II and V CTLD Zebrafish | xls | Materials and Methods, 15; Dicussion, 39 | file:///\\TS-QVHLC90\share\Erin C\Thesis Electronic Appendix\Group II and V CTLD Zebrafish.xls |
| TMHMM | File Folder | Materials and Methods, 15 | \\TS-QVHLC90\share\Erin C\Thesis Electronic Appendix |
| TMPRED | File Folder | Materials and Methods, 16 | \\TS-QVHLC90\share\Erin C\Thesis Electronic Appendix |
| FRR-pGEM-T Clone Sequences | File Folder | Materials and Methods, 19 | \\TS-QVHLC90\share\Erin C\Thesis Electronic Appendix\DNA Sequencing |
| FRR-pSecTag2 Clone Sequences | File Folder | Materials and Methods, 19 | \\TS-QVHLC90\share\Erin C\Thesis Electronic Appendix\DNA Sequencing |
| Older FRR-pSecTag2 Clone Sequences* | File Folder | N/A | \\TS-QVHLC90\share\Erin C\Thesis Electronic Appendix\DNA Sequencing |
| Other Sequences* | File Folder | N/A | \\TS-QVHLC90\share\Erin C\Thesis Electronic Appendix\DNA Sequencing |
| Erin's Freezer Catalogue | xlsx | N/A | \\TS-QVHLC90\share\Erin C\Thesis Electronic Appendix |

*Additional files from previous experiments that did not yield adequate clones

# AUTHOR'S BIOGRAPHY

Erin is originally from Southern Maine, where she grew up on the Magic School Bus, Nancy Drew, and buying nearly everything in New Hampshire. In middle school, Erin was introduced to infectious diseases and genetics in a life sciences class and never looked back. She graduated with Honors from Marshwood High School in 2009 and came to the University of Maine soon afterward. Erin worked in Dr. Rob Wheeler's lab for almost three years at UMaine, including two summers. As an undergraduate, Erin presented her work at three different showcases and was the recipient of various INBRE fellowships through the Honors College, a Barry M. Goldwater Scholarship, and the Frederick H. Radke Memorial Award.

Erin intends to earn her Ph.D in Molecular and Cellular Biology through the University of Maine Graduate School of Biomedical Sciences and Engineering. Her academic experiences have inspired her to pursue a career as a research professor at a college or university. She intends to perform biomedical research that relates to immunology, hematology, and/or functional genomics. When not in the lab, Erin can often be found knitting, crocheting, playing her clarinet, reading, cooking vegetarian delights, or sipping various herbal teas.